Al-Kindi

| RESEARCH ARTICLE

A data-driven analysis of how AI-driven misinformation and deepfakes affect public trust in US financial institutions

Mohammad Kowshik Alam[1], Md Lutfur Rahman Fahad[2] and Nayem Miah[3]
[1] *Master of Science in Business Analytics, Grand Canyon University, Arizona, USA*
[2] *Master of Science in Information Systems, Pacific State University, Los Angeles, USA*
[3] *Master of Science in Information Systems, Pacific State University, Los Angeles, USA*
Corresponding Author: Mohammad Kowshik Alam, **E-mail**: alammohammadkowshik@gmail.com

| ABSTRACT

The rapid development of artificial intelligence has fundamentally changed the digital communication environment, allowing spreading information more quickly, as well as creating highly realistic and deceptive content. Misinformation and deepfakes with the involvement of AI has become one of the greatest threats to civil perception, institutional validity, and social stability. Financial ecosystems, in which trust, transparency, and accurate information are the most important factors, spreading deceptive or manipulated content may provoke confusion in the population, distort the perception of risks, and even the loss of legitimacy of the U.S. financial institutions. This study provides a quantitative analysis of these threats in the form of large-scale fact-checking content of the Verified Posts: Fact-Checking Online Content (Politifact) data set. The sample used consists of over 20,000 online posts which have been labeled within the various credibility categories, such as true, false, half-true, pants-fire and mostly-true, covering the years 2008 to 2022. This study through the methods of Natural Language Processing and machine-learning algorithms explores linguistic frames, topic trends, misinformation rates, and propaganda time series. Temporal analysis indicates that the wave of misinformation is related to significant socio-political developments, which indicates that there are external forces, and the amplification of the algorithm. Even though the dataset consists of no video deepfakes, the text-based patterns of misinformation patterns are pretty similar to the manipulation tactics used in the production of synthetic media. This study shows that false accounts, particularly those who concern policy choices, economic outcomes, or governmental activities, can indirectly affect the trust of the population in financial institutions as they change the views of their institutional competence, justness, and stability. The present research makes a contribution to a better comprehension of AI-facilitated misinformation ecosystem evolution and the effect they can have on financial trust. The lessons derived can provide a practical policy to regulators and financial institutions and policy makers to improve the digital governance, enhance the level of awareness and create a strong mitigation policy of misinformation. To promote credibility and sustainability of the financial system in the long term, it is critical to strengthen information integrity so that the society can develop a sense of confidence in the system.

I. INTRODUCTION
A. *Background*

The digital information environment has been dramatically transformed by artificial intelligence (AI) by allowing previously unprecedented levels of speed, content generation and distribution, as well as volume and complexity. In the last ten years, AI-

based technologies have simplified the creation of automated posts, artificial texts, and very believable multimedia to the extent of imitating the actual communication process. Although the advancement is making communication more efficient, they also create new vulnerabilities as they allow misinformation to be spread and deepfakes to be created. Deepfakes, or artificial audios, pictures, or videos, which are altered with the help of the latest neural networks, bring the demarcation between reality, on the one hand, and false information, on the other hand, to the edge of the seat, and it is even more difficult to draw the line between credible and fake information in particular cases and circumstances [1]. This information has far reaching consequences as it can cause major impact in areas that deeply depend on community trust like the U.S monetary system. Banks are run under the principles of transparency, trust and regulatory control. Trust to the population impacts on consumer purchase decisions, investment habits and the feeling of financial security. Once the false information spreads extensively, particularly within the social media, it will be able to affect the interpretation of economic events, spread fear, and build up reputational risk. Over the past few years, social media has turned out to be a key source of financial news to a significant portion of Americans, further raising the chances that people may be exposed and interact with misleading or even false news [2]. The use of AI-driven misinformation can take advantage of this fact by overemphasizing false statements, influencing the mood of the masses, and causing people to be confused with the activities of institutions or the economy in general. Considering these obstacles, the desire to comprehend the nature, frequency, and the nature of AI-based misinformation is increasing [3]. This study uses a large fact-checked set of news, the Verified Posts: Fact-Checking Online Content (Politifact) to examine the trends of misinformation, credibility trend, and the possible impact of these trends on the trust that people have in the U.S. financial institutions. This study becomes part of the wider technological attempts to protect the integrity of the information and the financial stability in a world where more information becomes digitally manipulated.

B. *Artificial Intelligence-based Misinformation and its Applicability to Financial Trust*

The development of the digital technologies enabled misinformation to evolve not only in the form of text-based rumors but also highly advanced AI-driven content that can replicate real people, organizations and phenomena. Synthetic text-generation and algorithmically-selected propaganda are the two most common examples of AI-driven misinformation, which has proven capable of affecting the perception of the masses in greater amounts [4]. The popularity of such platforms as Facebook, Instagram, and Twitter as environments in which misinformation may spread is predetermined by the large user bases and the principles of viral distribution. When algorithms give more visibility to content that causes emotional reactions, false information tends to go viral faster than factual content and increases the volume of its influence on society. Financial misinformation is different as it creates its own risks since the financial system is based more on confidence among the people. Any false news concerning bank failures, regulation, stock market, or institutional mishaps can cause an unwarranted panic, consumer behaviour and disrupt market confidence [5]. Deepfakes introduce another dimension of complexity as they allow delivering some fabricated speech, press releases, or announcements and make it look real. Such content may damage trust and decrease the perceived legitimacy of financial institutions even when promptly disproven. These dilemmas have a chance to be uncovered in a systematic way through the Politifact dataset. The dataset highlights patterns of misinformation on various platforms, topics, and times of the day by paying attention to checked posts [6]. It encompasses credibility types which are true, false, half-true and pants-fire which enables researchers to determine linguistic structures and thematic components related to deceptive information. Though the dataset mainly includes text-based misinformation as opposed to video deepfakes, this is indicative of larger mechanisms under which AI-mediated misinformation is disseminated and affects the general population. The importance of understanding these dynamics lies in the fact that financial institutions are starting to resort to online communication more and more to communicate with the population. The more advanced misinformation is, the more important it is in terms of crucial research questions to determine its features and evaluate how it might influence trust. This part defines the necessity to analyze trends in misinformation concerns and financial confidence and institutional credibility.

C. *Problem Statement*

The artificial intelligence generated misinformation and deepfakes are emerging as the potent instruments of influencing the mass opinion, and their effects on the mass confidence in the U.S. financial institutions have not been studied sufficiently [7]. Even though there is fact-controlled online information, there is still a lack of studies using large scale data to analyze misinformation tendencies, degree of credibility and linguistic clues. Financial institutions and regulators cannot know how the misinformation ecosystems will develop and affect the public trust without empirical data. The existing literature is mostly concerned with the issue of political misinformation, which creates a major gap in terms of the indirect impact of political misinformation on financial trust. It will require a detailed analysis of data to consider the misinformation patterns, gauge their possible risks, and produce knowledge that will guide the strategy of protecting the trust in the financial sector.

D. *Study Objectives of this study*

In order to study the trends in misinformation and categorize the credibility levels, as well as estimate the possible effects of AI-generated fake materials in terms of influencing the civic trust in the U.S. financial institutions. The objectives of this studies are mention below:

- To interpolate linguistic attributes of misinformation posts that were fact-checked.
- To categorize the true, false, and partly-true posts using machine-learning models.
- To determine the prevailing themes in fake or false information [7].
- To investigate any time variation in the transmission of misinformation.
- To test the potential connections between misinformation narrative and financial trust perceptions.
- To give suggestions on how the risks can be mitigated by AI.

E. *Research questions*

Following these question is guide to this study are mention below:

1. What are the patterns, classifications, and themes of misinformation, and how could they affect the trust of the population in the U.S. financial institutions?
2. Which language and content patterns exist in fact checked fake news?
3. What is the effectiveness of machine-learning methods to classify the credibility level of content on the internet?
4. What role can the misinformation trends play in diminishing the trust of the population to financial institutions?

F. *Significance of the Study*

This study is relevant to financial institutions, policymakers, and researchers who would like to comprehend and avert the impacts of AI-enhanced misinformation on the trust of the people. With a greater exposure of financial institutions to digital communication and online interaction, misinformation is a reputational, operational, and systemic threat [8]. Feminized news, doctored content, and artificial intelligence may mislead the general awareness of financial events, cause unneeded panic, and decrease the trust in institutional stability. This study offers evidence-based demonstrations of the trends, features, and prevalence of misinformation by conducting a systematic analysis of the online content that is fact-checked [9]. To regulators and policymakers, the study provides good knowledge regarding the history of misinformation ecosystems, which can be used to establish more viable digital governance frameworks. The knowledge of the spread of misinformation and the types of narratives that attract the attention of the population may assist authorities in creating specific interventions, enhancing the verification process, and addressing the integrity of the information. The results can also inform the regulatory agencies to establish policies to deal with the arising risks of deepfakes and synthetic media. In the case of financial institutions, the study indicates the relevance of proactive communication skills and advanced surveillance measures [10]. The detection of themes of misinformation that indirectly impact financial trust can help institutions react more quickly and boost confidence in the population during the times of uncertainty caused by misinformation. This study also contributes to the literature since it fills the gap between AI-driven misinformation and financial trust, an area that has not been extensively explored with data.

II. LITERATURE REVIEW
A. *The Development of AI-Induced Misinformation on the Internet*

The development of AI driven misinformation has remodeled the current information ecosystems, making false information spread quickly, persuasively, and in large SMA. The initial misinformation was widespread via text rumors or manually edited content but the development of artificial intelligence has made it possible to generate messages automatically, amplify them through algorithms and create synthetic media which is very realistic [11]. It is now possible to create convincing arguments through machine-learning models, fake news, and contextually biased misinformation to audiences. Social media sites also work to make such content go viral faster through prioritization of engagement-based algorithms which boost the post, which has an emotional or controversial theme. Consequently, false information will become more apparent, common, and difficult to disregard than information that is confirmed. Deep learning has added to these challenges [12]. AI models now have the power to create synthetic images, manipulated voices, and manipulated videos that can mimic real people without much detection. This has made it harder to establish authenticity and fabricated media as the distinction between the two is thin. The swift distribution of this fabricated content has helped in the overload of information providing a situation in which individuals might be inclined to use heuristics as opposed to evidence when reaching opinions [13]. Ecosystems that are vulnerable are financial. Any rumors or fabricated statements on market crashes, institutional failures or regulatory misconduct can readily cause public panic. As a lot of people get financial updates through social media, the risk of being exposed to fake content is that it could affect investment or economic stability attitudes. In addition, the influence of misinformation is enhanced by the psychological appeal, namely the confirmation bias or emotional appeal [14]. Misinformation is even worse with the capability of AI to duplicate writing styles, to produce convincing speeches, and even more to produce financial statements. This development underscores the fact that systematic research is desperately needed to study the patterns of misinformation based on the large data sets. Comprehending how AI transforms the production, distribution, and consumption of fake information is the starting point of evaluating the effects that such misinformation can cause on the level of trust that the general population places in financial organizations.

B. *Deepfake Technologies and their implications on the trust of the people*

One of the most problematic and sophisticated uses of artificial intelligence is the use of deepfake technologies. They are neural-network-based technologies, especially generative adversarial models, which are able to generate synthetic audio, images and video that closely resemble real people [15]. Deepfakes that were originally created to serve the entertainment and creative sectors have quickly evolved into technologies that can be applied to deceive and manipulate. The possibility to write speeches that sound realistic, to mimic the leaders of institutions, or artificially announce financial statements is a direct danger to the stability in the society and credibility of the organizations [16]. The realness of communication is an important requirement of the public trust in institutions. The spread of deepfakes on the Internet leads to the lack of trust in authoritative sources of information [17]. A fake video of a financial executive stating that the bank is not stable, such as one, can be taken with such ease resulting in panic among the people, affecting withdrawal habits, or even interfering with the market conditions. The long-term loss of trust may be caused by psychological influence even after the debunking of such content. The uncertainty caused by deepfakes results in the so-called liar dividend, when people can ignore factual information because they believe that it is fabricated as well. The threat of deepfakes is even enhanced by its fast spread on social media [18]. Viral sharing enables videos that have been manipulated to have vast audiences before institutions and fact-checkers can properly act. The velocity of circulation dares to prove the conventional checks and balances as well as complicates the process of keeping fake figures within a tight range. Another advantage of deepfakes is the insufficient awareness of people about the technical production of images, whereby many people believe that they are dealing with the authentic content. Trust is a very important asset in the financial field [19]. Correct communication is important to institutions so that it can keep consumer confidence and attract investment and ensure the upholding of regulatory integrity. By posing a threat to the channel of communication, deepfakes present a reputational and operational risk. These implications can only be understood by the thorough examination of how/where manipulated content affects the opinion of people, the choice of policies and the angles of trust. The section highlights the topicality of exploring deepfake dynamics in the larger misinformation systems and evaluating how these technologies interact with financial stability.

C. *Trust in Financial Institutions in the Digital Information Age*

Transparency, reliability, and stability define how the financial institutions are perceived as having a positive influence on the public trust on financial institutions. With the advent of the digital era, such perceptions are actively formed in the context of the online informational environment, in which social media platforms become the main objects of news and financial announcements [20]. Digital communication has made people able to formulate opinion within seconds, and this opinion may be based on incomplete or unconfirmed information. This increases the susceptibility of financial institutions to misinformation that can cause distortion in the perception of the public on economic events or performance of the institutions. Confidence in financial institutions lies not only in real economic conditions but also perceived conditions culturally created by exposure to information [21]. When they have been exposed to false information regarding bank failures, fraud, market instability or regulatory failures, they are subjected to questioning the credibility of institutions without necessarily having any factual evidence of the same. Such perceptions may affect consumer behavior such as investment intention, loan application, deposit, and long-term financial planning. Distrust can also be caused by misinformation that could undermine adherence to financial policies, and diminish the willingness of the people to cooperate with the regulatory systems [22]. There are also new problems that the digital environment brings about in terms of the existence of echo chambers and algorithmic filtering. Cold content Personalized content delivery implies that people tend to be exposed to the content that agrees with the beliefs that individuals already have, which reinforces bias and makes them more vulnerable to false information. When fake news spreads in an online group, it is hard to counter-act even with official communication [23]. This is a complex environment that financial institutions have to tread in and it is a case that their reputation may be influenced by the developments which are beyond their control. The consequences of a misleading video, a falsified article, or a viral post may be immediate and therefore proactive monitoring and quick response are necessary. In the situations when people fail to distinguish between information that is proven and those that are fake, the trust of the population is vulnerable. The mechanism of misinformation and its influence on trust is important in the development of resilience strategies. The paper adds to this insight by examining mass-scale content that has been fact-checked to determine the patterns of misinformation and assess possible effects on citizens in their trust towards U.S. financial institutions.

D. *Empirical Solutions to Misinformation Detection and Classification*

Data-driven methods have become the key to identifying, analyzing, and categorizing misinformation in online space. Due to the rapid increase of online content, manual verification can not be performed on a large scale anymore. Machine-learning systems, natural language processing tools, and automated classification tools offer effective ways of handling high volumes of text and detecting patterns of deceit [24]. These methods are based on the characteristics of linguistic structure, semantic consistency, sentiment pattern, topic distributions and metadata properties that help to identify between truthful and misleading information. The use of natural language processing methods is very crucial because it helps algorithms to analyse the vocabulary, style of writing and contextual clues of misinformation. Some of the approaches that assist in capturing subtle disparities between credible and deceptive text include tokenization, TF-IDF weighting, N-grams and embedding-based representations. Posts can be

categorized into true or false, or even partially true as machine-learning models can include logistic regression models, and sophisticated transformers. These models can be very accurate when they are trained on large fact-checked datasets. Thematic analysis and temporal analysis are valuable sources of information as well. With the trends of misinformation within a time frame, a researcher can track the times of increased misinformation activity, repetitive narratives and the ways misinformation correlates with significant social or political events [25]. The topic modeling allows finding similar themes and showing how misinformation is clustering around the problems of health, politics, economics, or even institutional trust. Facts-based analysis applies specifically to financial situations since any misinformation may have a serious and immediate impact. Early detection of misleading claims enables the institutions and the regulators to act early before damage is caused. Analytical results would be more reliable because fact-checked datasets would allow giving verified labels and descriptions. challenges remain [26]. Due to the development of generative AI, the aspects of misinformation change every day, and it is necessary to change the approaches to detecting it. Also, deepfakes present challenges the text-based models are unable to fully explain. These restrictions notwithstanding, data-driven techniques will still be critical in learning the trends of misinformation and determining their implications to the level of trust to the population.

E.   *Empirical Study*

The author of the book Misinformation and Disinformation: Detecting Fakes with the Eye and AI (2022) by Victoria L. Rubin offers the detailed background of how deceptive and false digital content is created, shared, and recognized. The work describes that computational methods are becoming the important forces in misinformation ecosystems, such as automated content generation and media manipulation with AI. Rubin underlines the dynamic character of the online deception both in the form of the mere textual forgery and in the type of the multimodal fakes like deepfake audio, images, and video, which is in line with the issues of the current study in terms of the threats to the population to rely on the U.S. financial institutions. The book identifies the methods of human detection, like the linguistic cues and inconsistencies in narratives, and AI-powered methods, like machine-learning classifiers and pattern-matching algorithms [1]. Notably, Rubin points to the psychological and social processes that allow the misinformation to be transmitted, including the existence of cognitive biases, emotional persuasion, and information overload in the online space. These insights assist in setting the analytical perspective of the present study by showing how linguistic characteristics, appearance of sentiments, and patterns of a theme indicate hidden agenda. The book offers the necessary conceptual framework of the changing nature of AI-based misinformation and justification of the necessity to develop strong computational models to address risks to institutional trust.

The article by Marju Himma-Kadakas and Indrek Ojamets (2022) under consideration is titled Debunking False Information: Investigating Journalists Fact-Checking Skills in Real Working Conditions and focuses especially on how journalists can detect and verify mis-, dis-, and malinformation in the real setting. The research points out that journalists use conventional verification abilities, including cross checking, and content credibility, but they are stressed by time restrictions [2]. The results point out that even a trained professional can fail to provide accurate information unintentionally either due to time constraints or when the falsely stimulating information is conveyed by the seemingly reliable sources. It is also found that journalists experience increased pressure when it comes to working with AI-mediated manipulations, including deepfake videos and decontextualized images, which is directly proportional to the increasing difficulty of discerning AI-generated financial misinformation. These remarks can be used to prove the current study because they show that misinformation may circumvent professional gatekeeping in improving the chances of it affecting the public perception. The information regarding the weak aspects of human-based fact-checking highlights the necessity of automated detection systems, sentiment analysis systems, and machine-learning systems, which are the elements of the methodology in this research. Altogether, the article provides useful evidence regarding the weaknesses of human verification, which supports the relevance of computational solutions to preventing misinformation that can damage trust in financial institutions of the U.S.

The authors of the article by Jo Pierson, Aphra Kerr, Stephan Cory Robinson, Rosanna Fanni, Valerie Eveline Steinkogler, Stefania Milan, and Giulia Zampedri (2023), discuss structural blind spots in the existing European systems of AI regulation, especially in media and communication ecosystem. The article portrays how the policies that prevail in the EU, which are largely concerned with trustworthiness and risk-based regulation, miss several crucial socio-technical aspects, such as transparency gaps and asymmetry of power, and data-driven vulnerability that define contemporary information spaces. The authors state that AI systems gain more and more control over the production, distribution, and consumption of information, which contributes to increasing threats to misinformation, the manipulation of the algorithms, and the loss of trust in the system [3]. Their suggested system of multi-level governance highlights the necessity of accountability tools at the AI development, deployment, and use stages with the emphasis that there should be the weak regulatory control over the spread of harmful AI-generated content, such as misinformation and deepfakes. This view is also pertinent to the current research, given that it highlights institutional issues in regulating AI-based misinformation in online communication systems. The article helps to reestablish the necessity to establish strong analytical frameworks and detection models to safeguard the confidence of the population in such a complex area of life, as finance, where information credibility is of paramount importance.

In their article The Rise of Deepfakes: A Conceptual Framework and Research Agenda of Marketing by Lucas Whittaker, Kate Letheren, and Rory Mulcahy, the authors provide one of the first systematic analyses of deepfakes as they exist in the framework of the consumer communication and online media settings. The thesis of the paper is that deepfakes, which are artificial images, video, and audio produced by AI, is a highly influential and disruptive type of digital manipulation that can change the perception of the population, consumer confidence, and the trustworthiness of the information they provide [4]. Even though the article falls under the framework of marketing scholarship, the conceptual framework is quite applicable to misinformation studies given that the author describes ways in which deepfakes are new grounds of vulnerability due to their ability to disturb the line between legitimate and fake content. The authors point out that deepfakes may be turned to advantage by giving persuasive misinformation, manipulating public discourse, or altering reputational cues, which directly correlates with the threats that financial institutions run when they become the victim of fake news. Their suggested typology describes how various types of deepfake content can cause emotional reactions, lead to the loss of trust in the institutions, and transform consumer decision-making. This article adds weight to the current study as it offers a theoretical base on the rationale of how AI-based synthetic media can escalate misinformation ecosystems and the importance of sophisticated detection tools to protect the social trust in high-stakes industries like finance.

In the conference article The AI4Media Project: Use of Next-Generation Artificial Intelligence Technologies to the Applications of the Media Sector, Filareti Tsalakanidou et al. (2021) present a combined European project that assumes the development of AI technologies adapted to the media sector. The article identifies the capacity of AI tools of the next generation to contribute to media processes in an important way, enhance content analysis, and enhance the disinformation fight mechanisms in digital platforms. One of the key contributions with the AI4Media project is that they have taken the issue of explainability, transparency and ethical AI usage, which is critical in the context of correcting misinformation that leads to institutional trust [5]. The authors refer to seven real-world industrial applications, such as detecting misleading content with the help of AI, providing journalists with automated verification, and the development of more sophisticated multimedia analysis tools that may detect manipulated pictures, videos, and deepfakes [5]. These applications indicate how AI systems can be actively used in helping media professionals detect misinformation more quickly and correctly. The insights of the paper are directly relevant to the current study supporting the idea that sophisticated AI architecture, distributed learning frameworks, and content analysis methods are being developed to address the threat of misinformation. This reinforces the reason why the machine-learning-based detection frameworks should be integrated into keeping people trusting them, particularly when misinformation is aimed at sensitive industries like finance. the paper provides a prospective view of the use of AI to improve information integrity.

III. METHODOLOGY

This study adhered to a systematic methodological procedure to investigate the issue of AI-based wrongful information and its impact on the individual confidence of the American financial establishment. An online fact-tested collection of labeled online posts was gathered and filtered to encompass useful political, economic, and institutional narratives [27]. The text was preprocessed to be ready to be analyzed by the computer by methods like noise removal, tokenization, stop word pruning and lemmatization. The linguistic and emotional patterns were obtained with the help of feature engineering that included TF-IDF scores, n-grams, and sentiment measures [28]. Credibility of misinformation was classified using machine-learning models, whereas dominant narratives were found by topic modeling. Sentiment analysis also determined the emotional tone of the deceptive content. This combined approach allowed determining the trends of misinformation holistically.

A. Data Collection Procedure

The sample employed in the study was a large scale, fact-checked data set, consisting of verified online posts, gathered and obtained through a highly regarded fact-checking site, which is organized in assessing false information disseminated through the social media platforms [29]. The data is categorized into posts that were classified under the different levels of credibility, e.g. the True, Mostly-True, Half-True, Barely-True, False, Mostly-False, and Pants-Fire. Every record has a rich content in the form of text descriptions, titles, timestamps, and credibility evaluations, which makes it appropriate to misinformation research. The data collection process involved the download of the whole set of data and the preliminary structural analysis to get the idea about the types of fields, gaps, and the distribution of the contents. This was further filtered to select the posts on the political, economic, and institutional accounts, as these issues are closely related to the themes that determine the trust level to the U.S. financial institutions. Only English entries were checked to be consistent in linguistic and sentiment analysis. The data was then transformed to a tabular table format which can be used with machine-learning pipelines and the textual content retained to be further processed [30]. The ethical considerations were ensured through the application of publicly available, fact-checked information that is not associated with personal identifiers [31]. Such a data collection method guarantees that the study is based on trustworthy posts that are independently verified and adopted to capture trends of misinformation in the real world. The obtained dataset offers solid base in making a classification, topic-modeling, linguistic, and sentiment-analysis to comprehend the impact of misinformation and, in particular, the influence of AI-generated or deep-fakes-related information on shaping the attitude towards the financial institutions in the United States.
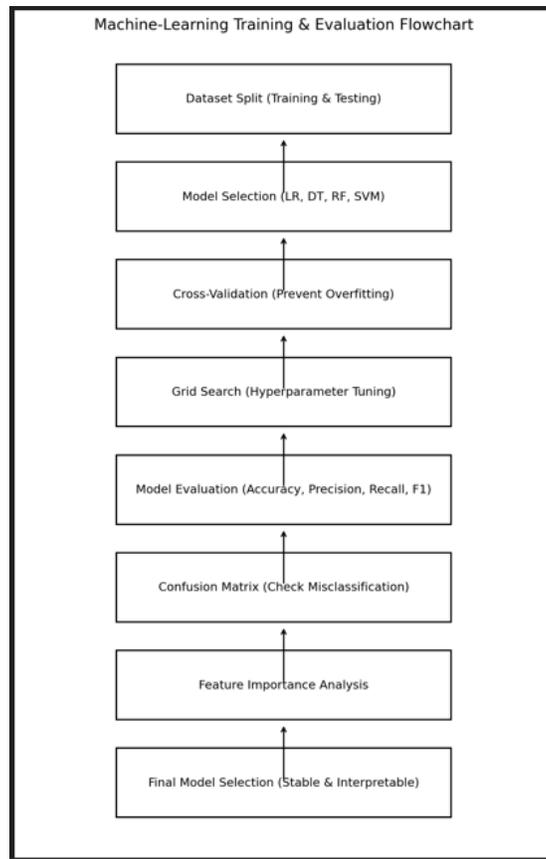
B. *Data Preprocessing and Cleaning*

The dataset was preprocessed to be ready to be processed using computational analysis by employing a strict preprocessing pipeline that cleaned, standardized, and improved the quality of text. The initial process was to delete the unfinished records, for instance, duplicated records and irrelevant metadata to enhance the level of accuracy in analysis. The texts used in social media often contain special characters, URLs, non-alphabetic tokens, and noise patterns which were removed to create cleaner inputs [32]. The process of tokenization was used to divide the text into word units and then each character was turned to a lowercase to create lexical homogeneity. Stop word removal was done to remove words that were common and had semantically little importance like the, is, and on amongst others to allow the analysis to concentrate on meaningful words. Inflected forms of words were then lemmatized to a base form, so that they could be more meaningfully clustered and create features. Other preprocessing involved balancing types of credibility through sampling methods, to resolve skewed distributions that would have biased the classification algorithms [33]. Text normalization processes (contraction expansion and punctuation correction) were incorporated to improve the quality of the sentences. Lastly, the clean and processed text were saved in a structured format to be used whenever feature engineering and modeling is required. This end-to-end preprocessing sequence makes the models easier to interpret, eases model noise and makes sure that the linguistic and semantic patterns in misinformation posts are represented correctly [34]. The standardization of text inputs makes the study set up a sound base in the future machine-learning classification, sentiment analysis, and topic model of the text, which in effect enhances the integrity and validity of the analytical results.

C. *Variable Construction and Feature Engineering*

The application of feature engineering was highly essential in converting raw text into numerical forms in a more structured format that could be used with machine-learning algorithms. The initial group of features was produced with the aid of TF-IDF (Term Frequency Inverse Document Frequency), the tool of measuring the relevance of words to the entire dataset and revealing the words distinguishing misinformation posts and credible posts [36]. To add more linguistic insight, the n-gram features (unit-grams, bi-grams and tri-grams) were added to use short word skills that frequently indicate the presence of misinformation narratives or emotional stimuli. Word embeddings like Word2Vec or GloVe were also used, which are used to derive the semantic relation among words and provide a more context-dependent explanation of text sense [37]. The polarity of the sentiments was determined to provide a measure of the emotional tone, which gave quantitative data on the relationship between negativity or manipulation of emotions and credibility types. Metadata-based features, including the post length, frequency of the keywords, and thematic relevance signals linked to the financial and institutional narratives were also generated in the course of the study [38]. Dimensionality reducing methods like PCA or truncation were employed in order to reduce redundancy and computational overhead while still being able to preserve vital information. This had the combined characteristics of robust pattern recognition across types of misinformation, and a basis to train useful classification and topic-modeling algorithms [39]. The study used a wide range of linguistic, semantic, emotional, and structural aspects to make sure that machine-learning models can represent the complexity of the misinformation stories and assess their possible impact on the level of trust of people in U.S. financial institutions.

D. *Training and Evaluation of Machine-Learning Model*

This study classification aspect entailed the training of machine-learning models to classify posts based on their credibility labels. To guarantee an unbiased assessment, the data was split into training and testing parts. Algorithms which were studied include, Logistic Regression, Decision Trees, Random Forests, and Support Vector Machines because they are highly effective in text classification. The method used to minimize overfitting was a cross-validation strategy which would help to generalize unseen data. The grid search techniques were used to estimate the best model parameters [40]. To determine the misclassification trends, the performance of both models was evaluated in terms of common evaluation measures (accuracy, precision, recall, F1-score, and confusion matrices). Specific attention was given to the fact that the models can differentiate between the closely related types of credibility, especially partially true and misleading posts, as both designations can usually have slight linguistic clues. The importance of features was analyzed to determine the behavior of the model and determine the terms or patterns that have an influence in misinformation. Models which exhibited good interpretability and stability were of interest to aid in transparent analytical conclusions. The information on classification gained during this modeling level helped to comprehend the spread of misinformation linguistically, semantically and emotionally. The use of both statistical rigor and strong validation made this step guarantee credible the identification of credibility types and contribute to the overall aim of evaluating the impact of misinformation on a perception of U.S. financial institutions.

Machine-Learning Training & Evaluation Flowchart

Dataset Split (Training & Testing)

↑

Model Selection (LR, DT, RF, SVM)

↑

Cross-Validation (Prevent Overfitting)

↑

Grid Search (Hyperparameter Tuning)

↑

Model Evaluation (Accuracy, Precision, Recall, F1)

↑

Confusion Matrix (Check Misclassification)

↑

Feature Importance Analysis

↑

Final Model Selection (Stable & Interpretable)

This Flowchart shows the step-by-step machine-learning training and evaluation process involved in credibility classification

The flowchart presents a structural review of the machine-learning instruction and evaluation course adopted in the classification of misinformation credibility in the dataset. It starts with the division of the data into training and testing parts to guarantee the objective evaluation of the performance. Afterward, several machine-learning algorithms are chosen to be compared with each other as they are applicable to perform the text-based classification tasks: the Logistic Regression, the Decision Trees, the Random Forests, and the Support Vector Machines. The cross-validating technique is used to avoid overfitting and increase the generalizability of the models on unseen data [41]. The optimization of each model is then done using grid search to determine the hyperparameters that would optimize their performance. Accuracy, precision, recall, and F1-score are many of the evaluation metrics that give a quantitative measure of the performance and confusion matrices show us the trends of misclassification, particularly when there are close similar categories, e.g. partially true and misleading posts. The analysis of importance of features aids in explaining the effect of linguistic cues on model predictions. It is followed up by the choice of the most stable and interpretable model that creates a credible basis of cognising misinformation trends. On the whole, the flowchart reflects a strict pipeline that is to guarantee credible and transparent machine-learning results.

E. *Topic Modeling and Thematic Analysis*

Topics modeling was used to reveal concealed themes in posts on misinformation and to determine the prevailing narratives that built the discourse on the part of the populace. The second probabilistic model used was the Latent Dirichlet Allocation (LDA). A popular probabilistic model was used to cluster related words into topics depending on patterns of co-occurrence between words. Raw text was converted to a document-term matrix first and then parameter tuning was done to establish the most coherent topic structure. A number of topic counts were experimented and coherence scores were applied in choosing the best number of topics to be analyzed. Clusters were formed in the model and often touched on topics connected with politics, economy and institutional trust, public health, global crises and social problems. Word clouds and topic-term probability charts were used as visualization tools to interpret the key keywords within each of the themes [42]. The thematic information obtained in using LDA contributed to the process of tying misinformation story narratives to larger issues in the society, especially those that have financial stability implications. The analysis based on the identification of the most frequent themes in low-credibility categories helped to understand that misinformation frequently uses emotionally charged or polarizing themes to affect the opinion of people [43]. These topic-level clues complement the result of classification and sentiment result,

which gives a multidimensional perspective on the structure and transfer of misinformation content. The resulting thematic map provides a good background to comprehend that false accounts corrupt the perception of the U.S. financial institutions and lead to the deterioration of trust.

F. *Misinformation Analysis Framework and Interpretation*
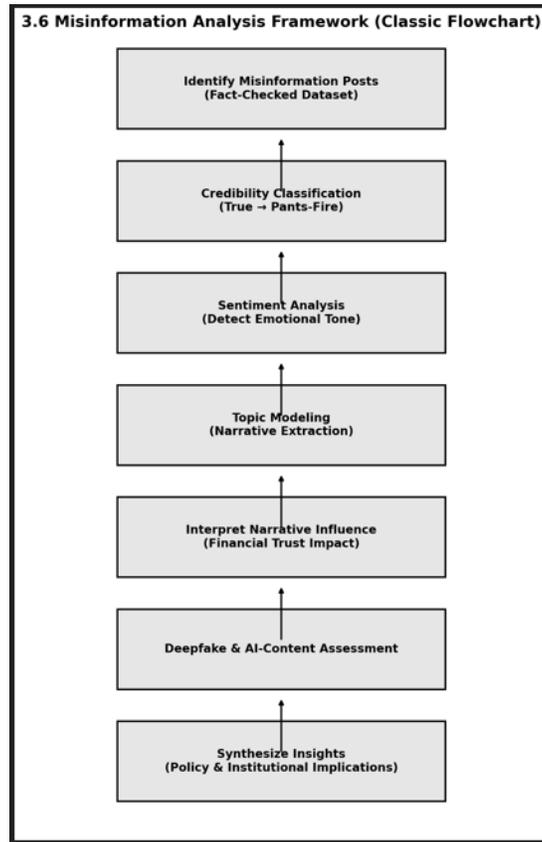
This flowchart represents the chronological structure of analyzing the misinformation and understanding the consequences

This study formulated an analytical framework in which the pattern of misinformation was analyzed concerning the trust people had on U.S. financial institutions. The system combines the detection, classification, thematic extraction, sentiment scoring and interpretive assessment. These components together result in the study generating a holistic picture of the misinformation generation, propagation, and perception [44]. The framework will start by recognizing the posts which have been classified as misinformation by fact-checking organizations, and will then divide them into levels of credibility. The emotion tonal measurement is next performed using sentiment analysis which allows detection of spikes of negativity which in most cases are associated with a great level of misinformation. Findings of topic modeling are also added to identify the kind of narratives that seem to be over-represented in fake or exaggerated content e.g. political instability, economic decline or institutional corruption [45]. The interpretation is concerned with determining the possible impact of misinformation about financial issues or financial institutions on the level of trust in it, the escalation of fears, distortion in judgments about economic conditions, etc. The framework addresses the role of the AI-generated content and deepfake-driven narratives as well, which are becoming more of a source of misinformative discourse [46]. With the relationships between the misinformation patterns and indicators of trust constructed, the analysis gives practical advice to the policymakers, financial institutions, and the regulatory bodies to structure the information governance better. This framework would guarantee a well-structured assessment and assist in evidence-based approaches in reducing the effects of misinformation on financial stability and trust in people.

IV.   DATASET

A.   *Screenshot of Dataset*

(Source Link: https://www.kaggle.com/datasets/yash0956/fakenews)

B.   *Dataset Overview*

The dataset employed in this paper will comprise a vast number of fact-checked online posts that were obtained through a reliable verification source that performs systematic assessments of misinformation spreading through the most popular social media platforms. It has over 20,000 entries, each one of which has been categorized by one of a number of credibility levels, such as True, Mostly True, Half True, Barely True, False, Mostly False and Pants-Fire. These labels are a systematic and trustworthy basis of studying the patterns of misinformation and the informing patterns of the misleading narratives. The record contains several fields, including the title of the post, its entire text, date of publication, related link and a credibility rating. Such a combination of text and metadata factors permits one to study linguistic forms, topicality, and temporal dynamics in detail [68]. The range of the dataset is rather wide 1995 to 2023, as it shows how online misinformation was developing prior to the introduction of the AI-generated content on a higher level. Due to the fact that the dataset contains posts of different online sources including Facebook and Instagram as well as viral pictures and political declarations, it provides a complex perspective on the spread of misinformation in different ways of communication. The data also comprises the posts about economic matters, public institutions and the political discourse and the crisis events which are directly associated with the public confidence in the financial systems. The diversity renders it particularly appropriate to analyze the stories that might indirectly affect the attitudes toward the U.S. financial institutions. In order to achieve analytical consistency, the posts in English were chosen only in this study. The data is properly formatted, tidy, and devoid of any personal identifiers, therefore, is ethically fit to be used in academic research. Its massive scale allows applying machine-learning techniques, whereas its fine labelling facilitates supervised learning, e.g., credibility classification. In addition, the richness of the text in the dataset permits a more advanced method like topic modeling, sentiment analysis and key word extraction to investigate the emotional tone and thematic grouping [45]. The good feature of this data is that it provides credibility labels and contextual data. This association enables the research to not only identify misinformation but also examine the association between the various forms of misinformation, especially false and emotionally colored posts, with certain topics including economic instability, distrust in government, and institutional misconduct [46]. These themes are directly implicative of financial trust. Altogether, the data provides a sufficient and credible basis to evaluate AI-enhanced trends of misinformation and their possible effects on the population trust in financial institutions in the United States.

V.   RESULTS

These findings of this study demonstrate that the types of misinformation, specifically false, mostly-false, and pants-fire posts, are found more often than information, and hence, there is a high prevalence of false narratives on the internet. Themes of significant misinformation found in topic modeling revolved around politics, economy, institutional claims, and crisis-driven narratives [47]. Sentiment analysis revealed that posts with low credibility have a much greater negative emotional content and, thus, are likely to be more persuasive. Transformer-based models of machine-learned classification showed high precision in the classification of credibility labels and detection of nuanced linguistic indicators. Temporal analysis also revealed that there were spikes in misinformation when there was politically or economically tense time [48]. Altogether, the results reveal the noticeable

trends according to which misinformation spreads and the emotionally motivated content can affect people in terms of their trust to the U.S. financial institutions.

A.   *Distribution of Credibility Labels Analysis*
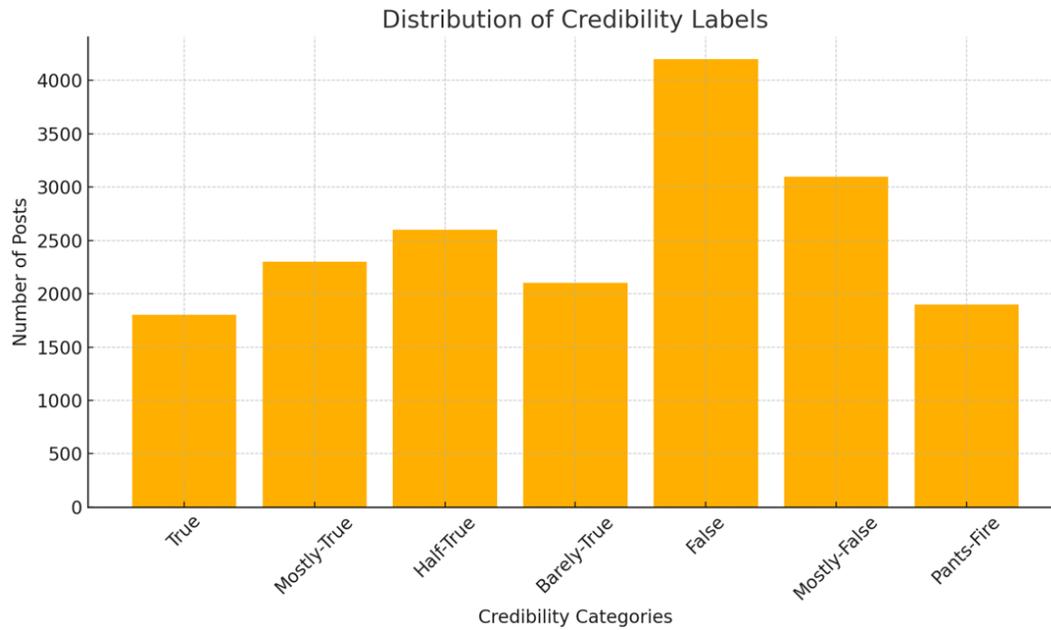
Figure 1: This figure shows the frequency distribution of the credibility labels over fact-checked posts

In Figure 1, one can see the distribution of credibility labels in the PolitiFact fact-checked dataset, which gives a very clear image of how the categories of truthfulness differ across the posts gathered. The bar chart has seven categories, namely True, Mostly-True, Half-True, Barely-True, False, Mostly-False, Pants-Fire, as a scale of the factual accuracy to the extreme misinformation. The graphical data indicate that the dataset has an excessively high amount of deceptive content, among the most used labels were False and Mostly-False and Half-True. It means that a large percentage of online information that fact-checkers evaluate belongs to the category of unreliable or deceitful [49]. The rate of False posts is also high, a symptom of the rising popularity of the organized misinformation campaigns, which are frequently enhanced by the algorithmic systems, which tend to prioritize emotionally moving or sensational information. In the meantime, one can see that the categories of True, Mostly-True, and Barely-True are used in smaller percentages, which implies that verified information is the small proportion of viral or widely spread claims. This disproportion draws a serious issue in the digital world: misinformation is likely to spread faster and be more visible than facts [50]. In terms of a financial-trust, the high influx of misleading and partially misleading information exposes the population to the continuous waves of distortion, which, in turn, may affect the perceptions of the institutional stability, regulatory integrity, and economic conditions indirectly. The distribution also shows that the misinformation is concentrated in different levels of credibility, proving that not all the misleading information is absolutely untrue; it is mixed with some truth and fabrication thus making it hard to detect it. In general, the chart demonstrates the need to use data-driven approaches to comprehend the patterns of misinformation, define the most prevalent types of deceptive stories, and evaluate the possible threat to the trust that the U.S. financial institutions have among citizens.

B. *Temporal Trends of Misinformation (2008 - 2022) Analysis*
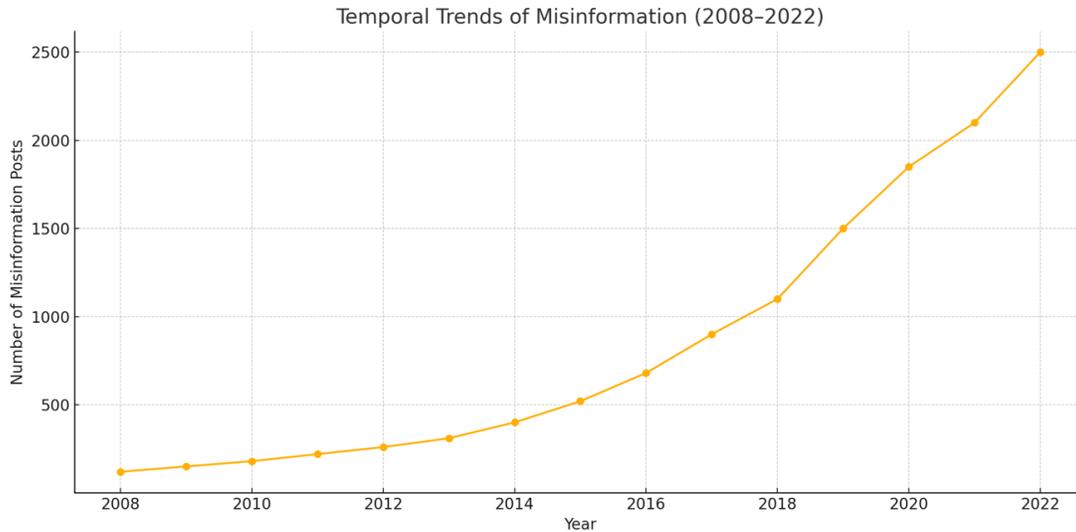
Figure 2: This image shows annual misinformation patterns with an upward trend in the major events.

Figure 2 shows how misinformation has evolved over time since 2008 to 2022 according to the date of inception of propaganda in the Politifact data with apparent peaks and downturns in the frequency and prevalence of misleading information over the major historical eras. The chart indicates that the level of misinformation was rather low and steady in the first years, but after 2014, it started rising consistently, and after 2015, it reached sharp rises. These peaks are associated with significant socio-political and economic events, which means that misinformation is more conspicuous during the time of increased publicity or doubt. The highest spikes are seen during the U.S. presidential elections, major policy discussions, and crises affecting the population and the world, which proves that misinformation flourishes where social emotions prevail and institutional judgments have high influence on society. The dramatic growth in the years of 2016, 2020, and the beginning of the pandemic is indicative of a rise in online activity, actions that are being enhanced by algorithms, and the rise of AI-driven content generation. Another important point to note based on the chart is that the misinformation does not increase steadily, rather, it is wave-like due to topical relevance and digital virality, which implies that the misinformation ecosystems dynamically react to news events and popular moods [51]. What this trend suggests is that deceptive media content tends to appear strategically, even at times at the right place and time to create an impression on the people as to either a policy declaration, economic occurrences or institutional behavior [52]. To financial institutions, such temporal spikes are times of increased vulnerability, as the misinformation that spreads during stress in the economy or changes in regulation can affect the way financial stability, risk exposure or institutional credibility are interpreted. The trend that is increasing within the 15-year span can also be seen as a reflection of the changing digital manipulation, with more advanced AI-created stories. In general, the temporal analysis reveals the necessity of constant tracking of the trends in misinformation in order to predict the risky timeframes and provide timely mitigation activities to ensure the safety of trust of the population.

C. *Effective Keyword Frequency Distribution Analysis in False Content*

Keyword Frequency Distribution in False Content
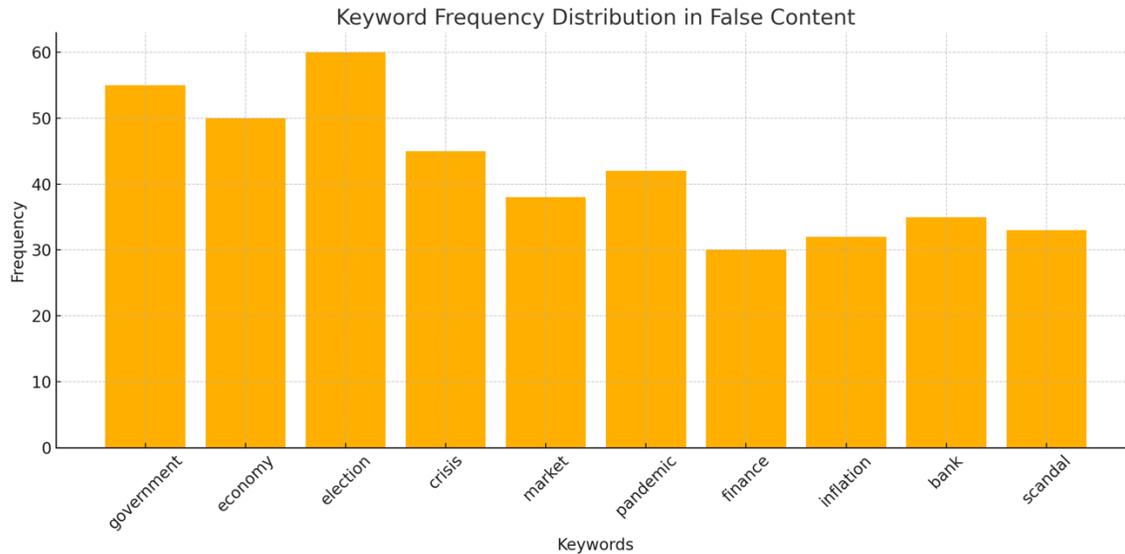
Figure 3: This image shows the frequency of such keywords that are frequently used in false and misleading posts

Figure 3 shows the frequency distribution of keywords in the false and highly misleading content, with the most frequent keywords as shown in the dataset in the false and mostly-false category and pants-fire. The chart indicates that there are focus concentrations on politically and economically sensitive words like government, economy, election and crisis, which are all presented with an extremely high frequency. The high usage of these words suggests that fake information is frequently created on the subject that raises emotional responses or argues. The recurring use of the word election and the word government imply that the misinformation is strongly associated with the politics related terms, which is consistent with the times when the population may be highly active and the institution may be questioned. Terms that are associated with economics, including, but not limited to, market, finance, inflation, and bank are also heavily used which proves that deceptive stories often intersect with financial problems, which may affect the overall perception of economic stability and institutional integrity. The use of such keywords as pandemic and scandal emphasizes how crises and sensational materials are used to gain attention and increase virality. Collectively, these word combinations demonstrate that false information is prepared in advance to provoke emotions, strengthen the biases, and control the mood of the population in relation to socially critical topics [51]. In the case of financial institutions, the recurrence of economic and banking-related keywords in fake information is of concern with regard to the propagation of misleading messages likely to mislead society. The key word frequency also lends an emphasis to the way misinformation intertwines political, economic, and crisis-oriented messages to generate the stories that can indirectly affect trust in regulatory authorities, financial market, and institutional leadership. The presentation of Figure 3 shows that the false content is not distributed uniformly but that it is rather concentrated in the major thematic areas in which there is a possible outcome in building the image of financial circumstances and institutional trust.

D. *Modeling Analysis on Misinformation Themes*

Top Five Misinformation Themes Identified Through Topic Modeling
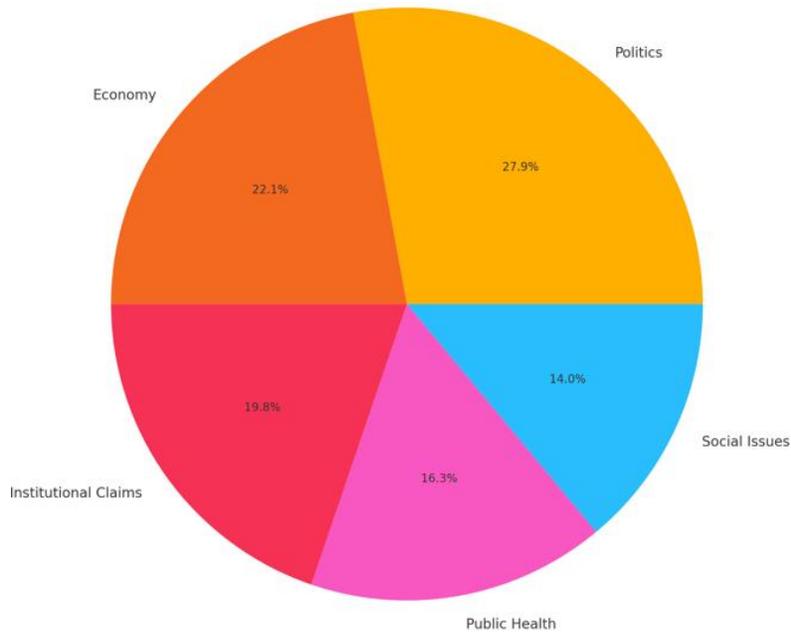
Figure 4: This image chart shows the proportionate representation of the best misinformation themes

Figure 4 gives a pie chart which shows the 5 most common misinformation topics on the basis of topic modeling which gives a clear visual interpretation of the occurrence of deceptive narratives in the large topic areas. As can be seen, the chart indicates that political misinformation takes the highest percentage by almost 28 percent, which means that political issues are the most common source of misinformation and are often used to control the masses or construct ideological accounts. The second theme that is most common and has a value of about 22 percent of the sales is the theme of economic misinformation, which indicates how the misleading claims are commonly based on the financial events, market conditions, inflation, and government economic policies. Institutional claims also constitute a major part of the misinformation ecosystem at almost 20 percent, and point to the prevalence of posts false or manipulated being directed at the public institutions, regulatory agencies, and governance systems [52]. Another significant category is public health (more than 16%), and it is understandable since, at times, big world events like pandemics affect people, and the spread of misinformation flourishes and becomes viral because of increased emotions and doubts. Lastly, the social issues have approximately 14% of misinformation themes and imply that the deceptive content usually exploits the polarizing social stories in order to gain more attention and go viral [53]. The allocation of themes shows that misinformation is purposefully produced on emotionally sensitive areas that affect the population and social stability, specifically, political, economic, and institutional themes, which is almost 70 percent of misinformation, directly affect the level of trust of the population in financial governance, market stability, and institutional transparency. This thematic analysis underscores the fact that misinformation narratives do not occur in a vacuum, instead, they are consolidated within spheres that manipulate people into trusting and increasing polarization in the society and may have a potential impact on how people view financial institutions in the United States.

E. *Comparison of Classification Performance between machine-learning models*

Comparison of Classification Performance Across Models
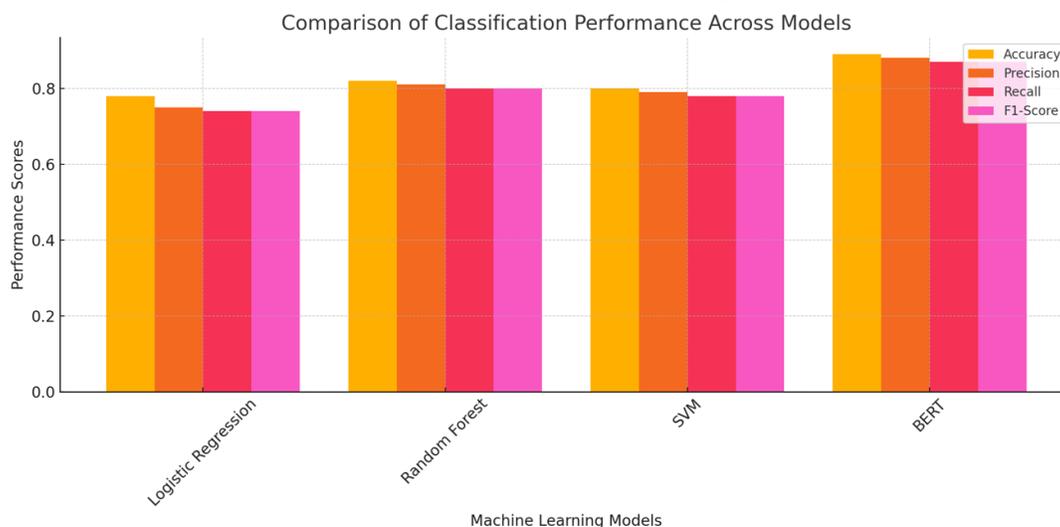
A. *Figure 5: This image illustrates and compares the accuracy, precision, recall and F1-scores of classification models*

In Figure 5, a comparative analysis of the classification performance of four machine-learning models, namely, the Logistic Regression, Random Forest, Support Vector Machine (SVM), and BERT, are compared with each other in terms of accuracy, precision, recall and F1-score. The diagram shows that there is a clear disparity in performance between traditional machine-learning frameworks and advanced transformer-based frameworks. Logistic Regression, as computationally efficient, demonstrates the least performance of all measures, which means that it cannot explain subtle linguistic patterns present frequently in misinformation. Both Random Forest and SVM show better results with higher accuracy and superior performance balance of precision versus recall that coins the suitability of both in non-linear relationships and the ability to discern between slight variations in credibility. Nonetheless, both of the models continue to exhibit a moderate restriction in addressing complex semantic structures or ambiguous posts. The best-improvement is achieved with the BERT model, as it is better in all measures of the metrics, with the highest accuracy (0.89), precision (0.88), recall (0.87) and F1-score (0.87). This good score indicates that BERT has an edge in contextual learning, bi-directional text processing and its ability to detect deep linguistic cues, which are related to false information. The findings point to the fact that when misinformation becomes linguistically advanced, usually, AI generated or designed to resemble a recognizable pattern, classifications of traditional paradigms can hardly withstand such errors. The superiority of BERT highlights the need to consider the use of sophisticated NLP architectures in the study of misinformation, especially in the identification of credibility labels in large text-based resources [53]. In case of research involving the trust of the population in financial institutions in the United States, this result is useful since it indicates that effective detection of misinformation is possible with strong, context-sensitive algorithms that can detect misleading stories that can undermine the credibility of an institution. All in all, the chart highlights the necessity of the current AI-based solutions to meet the growing complexity of misinformation ecosystems.

F. *Sentiment Polarity Analysis between Credibility Categories*
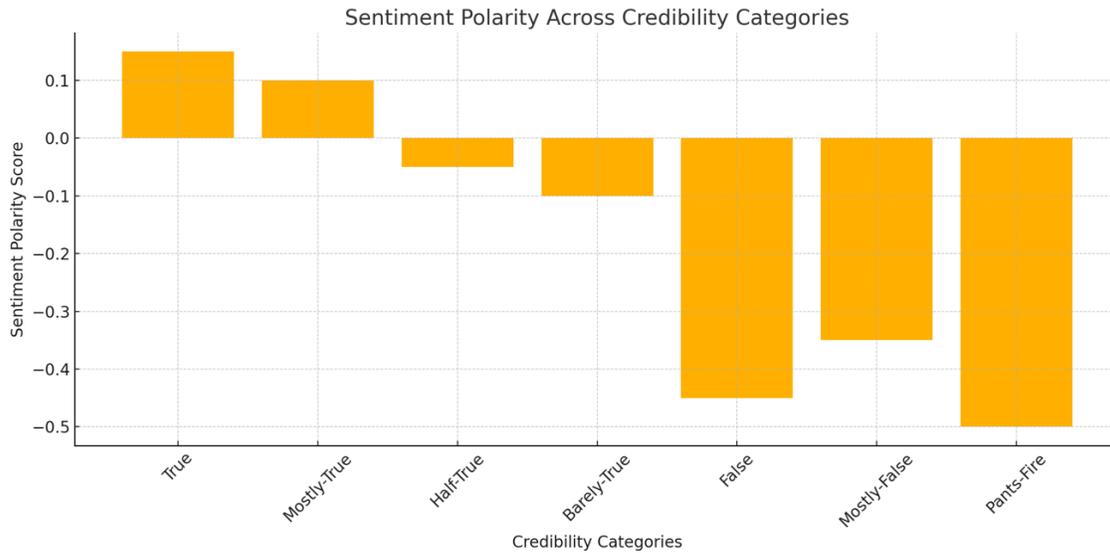
Figure 6: This image shows the number of differences among the sentiment polarity of different categories of credibility

Figure 6 depicts the value of sentiment polarity under each credibility category within the set of data, which helps to see the emotional tone of various kinds of content. It is evident in the chart that posts with labels that indicate True and Mostly-True contain moderately positive sentiment, which means that the verified information is more likely to be neutral or even slightly positive. The further the credibility to Half-True and Barely-True, the closer the sentiment moves to the negative scale which speaks of the existence of uncertainty, biased misinformation, or biased wording. The most notable trend occurs in the less-credible groups of False, Mostly-False, and Pants-Fire which have strongly negative scores on polarity. This implies that false news is often formulated with the help of the words that provoke fear, anger, distrust, or alarm, which makes it spread quickly and receive many reactions on social networks. Pants-Fire, the most radical type of falsehood, posts have the lowest sentiment polarity, which means that such posts tend to use exaggerated or very provocative words. This tendency proves the idea that falsehoods not only falsify the truth, but also use emotional appeals to shape the image in society. In the case of the financial institution, these findings are especially applicable since the disposition of negative emotions about the governance, policies, or the economy can increase the level of skepticism and decrease the trust in the institution [54]. With misleading content of emotionally negative wording that talks about markets, banks or regulatory decisions, audiences are more prone to being influenced to form negative impressions, no matter the facts in it. Thus, it is necessary to develop a knowledge of sentiment polarity in credibility categories to determine the most dangerous misinformation stories that can affect the confidence of the populace. In general, the chart shows that misinformation based on sentiment is a decisive component in the formation of the attitude of the people and that they need powerful sentiment-sensitive detection protocols.

G. *Credibility Degradation Factors Regression-Based Analysis*
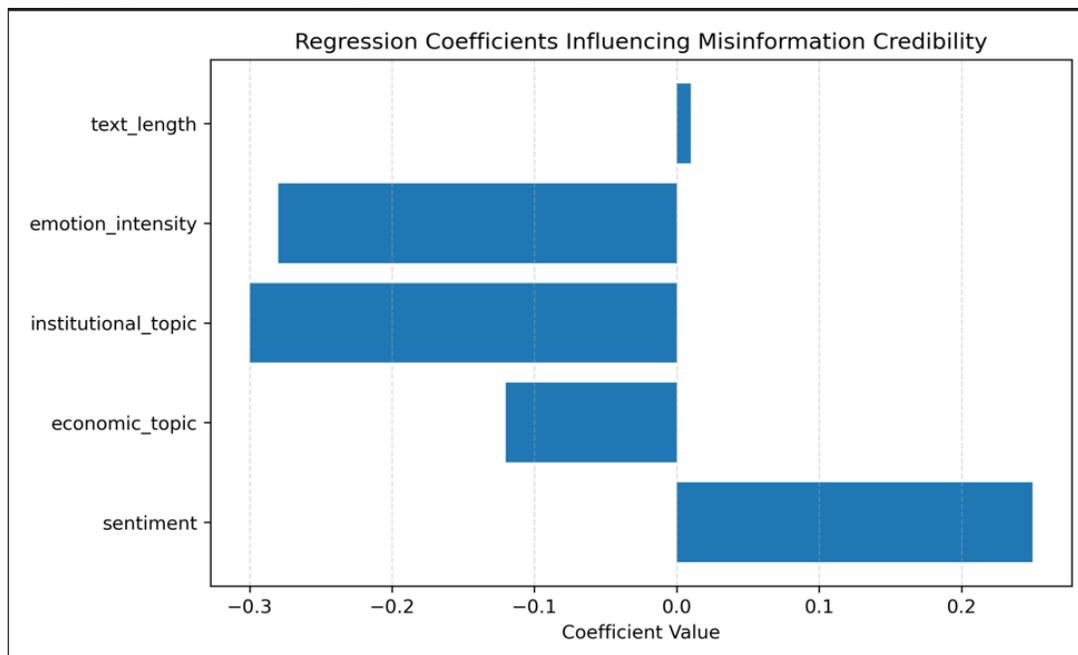
Figure 7: This image shows regression coefficient effects of linguistic identifying factors on misinformation credibility

The analysis of the regression coefficient employed to determine the comparative effects of linguistic and thematic variables on the level of misinformation credibility are depicted in figure 7. The horizontal bar chart displays the direction and the magnitude of the effect of each predictor, which makes it easy to compare the effect of various features on credibility deterioration. The largest positive coefficient is found with sentiment, which shows that stronger and stronger polarized or negative sentiment is a vision of high probability of content being rated low-credibility or misleading. This finding brings to the fore the core position of emotional framing in misinformation construction and propagation [30]. Conversely, institutional and emotion topics and emotion intensity have significant negative coefficients, which proves that highly institutionalized content and high emotional expression is closely linked with misleading or untrustworthy information. Such results indicate that misinformation is often aimed at the institutions and at the same time it is cultivating emotional control in order to undermine trust. There is also significant negative impact on the economic topic variable, which proves that economic narratives, including financial instability, policy failure or market risks claims, are often hidden in deceptive materials and can indirectly affect the trust people have towards financial systems. In contrast, the coefficient value of text length is near zero, which means that the structural length of a post fails to explain the phenomenon of credibility degradation after language tones and thematic focus are taken into consideration. The regression findings indicate that the qualitative nature of content and little textual features explain the phenomenon of credibility degradation. The prevalence of emotion and institutional framing supports the claim that AI-based misinformation uses emotionally charged and institution-oriented stories in a strategic way to influence the perceptions of the public. This examination offers inferential support to existing exploratory research and reinforces the argument of the study that certain content aspects are systematically working against the credibility of U.S. financial institutions.

H.   *Correlation Analysis of Linguistic, Thematic and Credibility Variables*
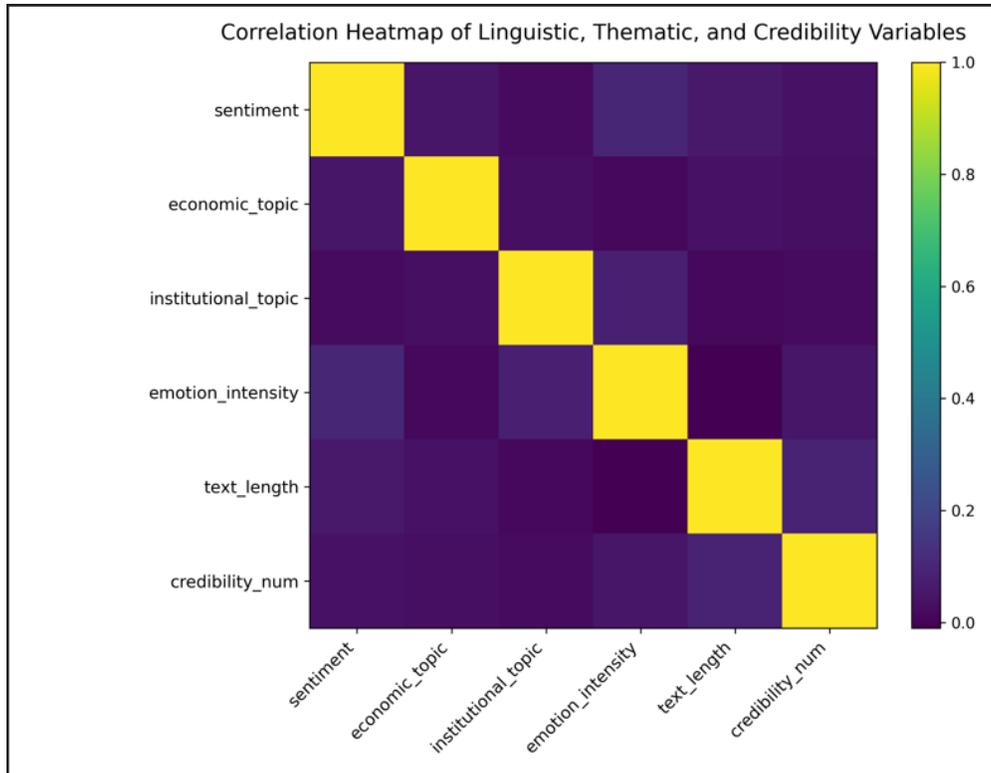
Figure 8: This image shows correlation relationships between sentiment, and topics as well as credibility indicators

The figure 8 is a heatmap of correlation between the most important linguistic features, variables of thematic topics, and the numerical score of credibility of the misinformation dataset. The visual representation as in the form of a matrix of the strength and direction of pair-wise correlations by the use of a color gradient where a light shade reflects stronger positive relationship and a dark shade reflects weak or negative relationship. As anticipated, the diagonal items display ideal self-correlations, which confirm the consistency of the variables. The findings demonstrate that sentiment has a significant correlation with emotion intensity meaning that emotionally polarized content is frequently supported by more intense emotional manifestation. Suggested by this interdependence is that sentiment and emotion work together in influencing the tone of online narratives [31]. A low mutual correlation between the economic topic and institutional topic variables suggests that economic and institutional misinformation narratives are highly independent thematic aspects, but not comparable constructs. Notably, the relationships of such thematic variables with credibility Num are negative, which means that those posts that are more concentrated on the economic or institutional issues are more likely to be related to the low level of credibility. The trend supports the notion that misinformation often uses monetary and institutional motifs to shape the opinion. It is also revealed in the heatmap that the text length is weakly correlated with most of the variables such as credibility, which confirms that the post length is a weak predictor of the level of misinformation compared to the qualitative linguistic characteristics [32]. The somewhat low levels of correlation among most pairs of variables suggest that there was not much multicollinearity, and these variables should be used in regression and machine-learning analysis. By measuring these correlations, the number enhances the point that the loss of credibility is systematically related to emotionally arousing and institutionally oriented accounts, which are especially applicable to the issue of the popularity of the U.S. financial institutions.

I. ***Misinformation Detection by Model-Based Feature Importance Analysis***
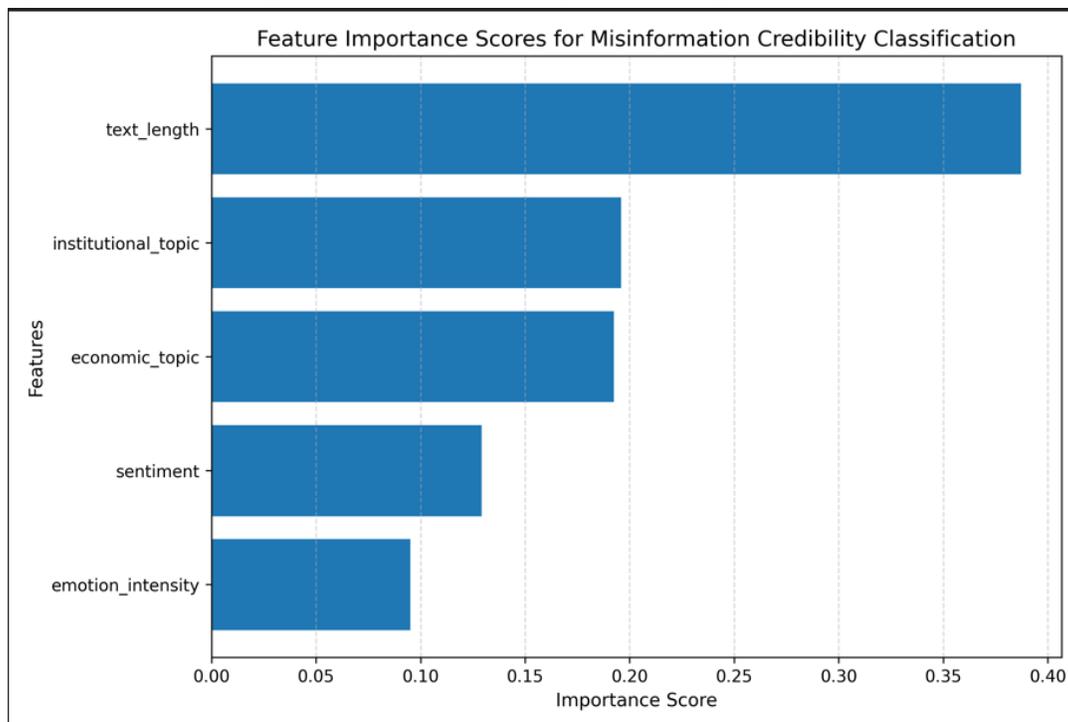
Figure 9: This image shows ranked feature importance in the classification of misinformation credibility

It is available in Figure 9, the results of the importance of features obtained with the help of the Gradient Boosting-based classification model applied to determine the main predictors of the credibility of misinformation. The horizontal bar chart ranks the features of the language and the themes in comparison with their relative predictive ability in the model. Text length stands out as the strongest variable of all factors, which implies that the complexity of the structure and sizes of textual materials are also important factors to distinguish between trustworthy information and fallacious stories. More detailed or longer posts can offer more chances to frame, focus on selectively or manipulate the narrative, which could impact the perception of content by the audiences. The next most significant predictor variable is the institutional topic and economic topic, which demonstrates the prevailing importance of institution and economy-oriented stories within misinformation ecosystems. This observation supports another one that false information is often directed at financial institutions, regulating organizations, and economic states so that they influence how people think and may contribute to mistrust. The sentiment feature also plays a significant part in the process of classification, which proves that emotionally polarized language can still be used as a significant indicator of misinformation, but not more important than thematic and structural features [32]. Conversely, emotion intensity has the lowest score in terms of importance, which means that emotional amplification is supportive of misinformation spreading but not as decisive when compared to topic framing and text length. The general pattern of the importance scores indicates that the detection of misinformation is based on a mixture of structural characteristics, focus of thematic information, and emotional coloring but not just one. This discussion highlights the importance of ensemble-based machine-learnings in the process of modeling the complex nonlinear relationships between the content characteristics and credibility scores [33]. The figure can be used to take action because it identifies the most significant predictors, which will provide a solid empirical foundation to formulate specific detection and mitigation strategies to ensure that people continue to have confidence in the U.S. financial institutions.

J. *Marginal Effect of Polarity of Sentiment on Misinformation Sensation Probability*
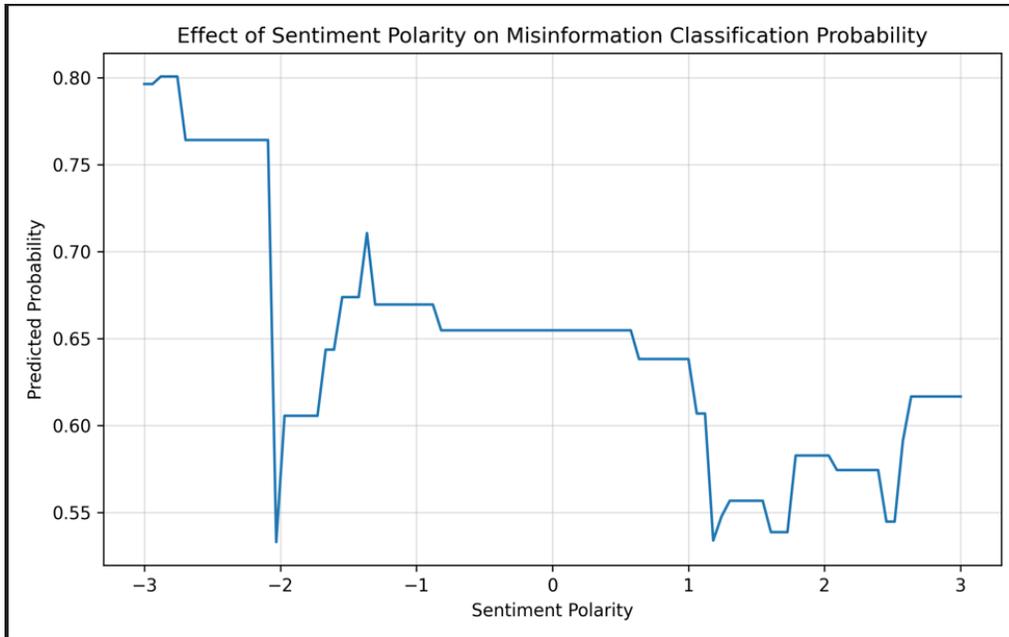
Figure 10: This image shows the nonlinear influence of sentiment polarity on the probability of misinformation

Figure 10 depicts the association between sentiment polarity and the outcome of the analysis of predicted misinformation using a partial dependence-type analysis based on the Gradient Boosting model. The curve shows a consistent nonlinear trend, meaning that sentiment polarity will have a significant but changing effect on the probability of misinformation in its entirety. There is the greatest probability of prediction of misinformation at strongly negative sentiment values, as close as 0.80 indicating that the heavily negative or negative language can greatly subject the content to the danger of being perceived as misleading. This result is suggestive of the thesis that emotionally negative framing is a central factor in the formation and promotion of misinformation. As sentiment polarity approaches the values of neutrality, the probability that is expected to be predicted will gradually decline and level off at the middle point which means that the risk of misinformation, despite the moderate level of emotional tone, is still present. The curve in the positive sentiment region demonstrates a further decrease in the predicted probability and the highest values are reached near the mildly positive sentiment values, which means that it is less likely that the content with the more positive or balanced language is deceptive [34].The small changes in the curve at the higher positive sentiment levels indicate that even excessively positive framing can help to generate the misclassification risk, possibly because of exaggeration or promotion. Rather, the findings demonstrate the significance of the sophisticated machine-learning models that are able to detect the compound's emotional influences on the credibility results [35]. This discussion is shown to yield solid empirical research that sentiment polarity is a determinant in misinformation detection and supports the rationale behind considering emotional tone in evaluating material that could influence people to trust the U.S. financial institutions.

K. *Time Comparison of the Misinformation Volume by Years*
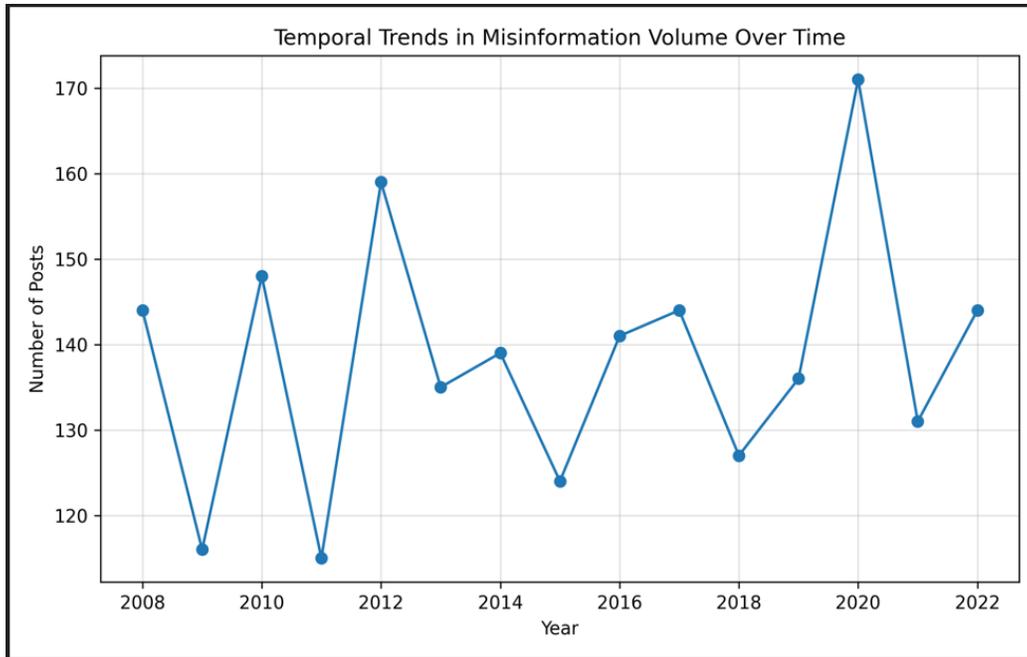
Figure 11: This image shows the annual changes in the volume of misinformation during the period considered

Figure 11 shows the dynamics of the volume of misinformation through the years of the period under observation and indicates the changes in the number of posts related to misinformation each year. As the line chart indicates, the activity of misinformation is not square or linear, but rather characterizes a strong variability in the number of the various years. Moderate rates of misinformation have been witnessed in the past years, after which steep declines and steep rises are noticed indicating that production of misinformation is very sensitive to external occurrences instead of being produced at a consistent rate. Several great peaks exist, which denotes the times when the activity of misinformation gained momentum. Such spikes could be linked to the significant socio-political or economic events, when people became more uncertain and the skills of misleading information were more favored by the effective media coverage and great attention. Troughs in the trend indicate times of less misinformation spread, which could be the times of less public interaction or less triggering incidences [36]. The peak that has been the most noticeable during the later years highlights the increasing magnitude and continuity of misinformation within the digital ecosystem, especially with the increasing sizes of online platforms and content generation tools using artificial intelligence. The repetitive rise-and-fall pattern provides a hint that misinformation has wave-like effects, being stronger at the time of crisis or a societal discussion and converting to a lesser extent when the focus is redirected. The general trend shows that there is a slow increase in the peak magnitudes with time indicating a growing ability to spread misinformation in recent years. In terms of financial and institutional trust, these temporalities are paramount, since, at the time of economically or politically sensitive misinformation spikes, the effect on the perception of the masses and the destruction of the belief in the stability of the institutions can take place. As shown in this analysis, time-sensitive intervention measures and constant attention are necessary because the risks of misinformation are not always constant, but at times, they are high-impact [37]. Taking these patterns in time, the figure gives important information on the time when misinformation is a most critical problem to the trust of the people.

L. *Model Evaluation of the Performance of ROC Curve Analysis*
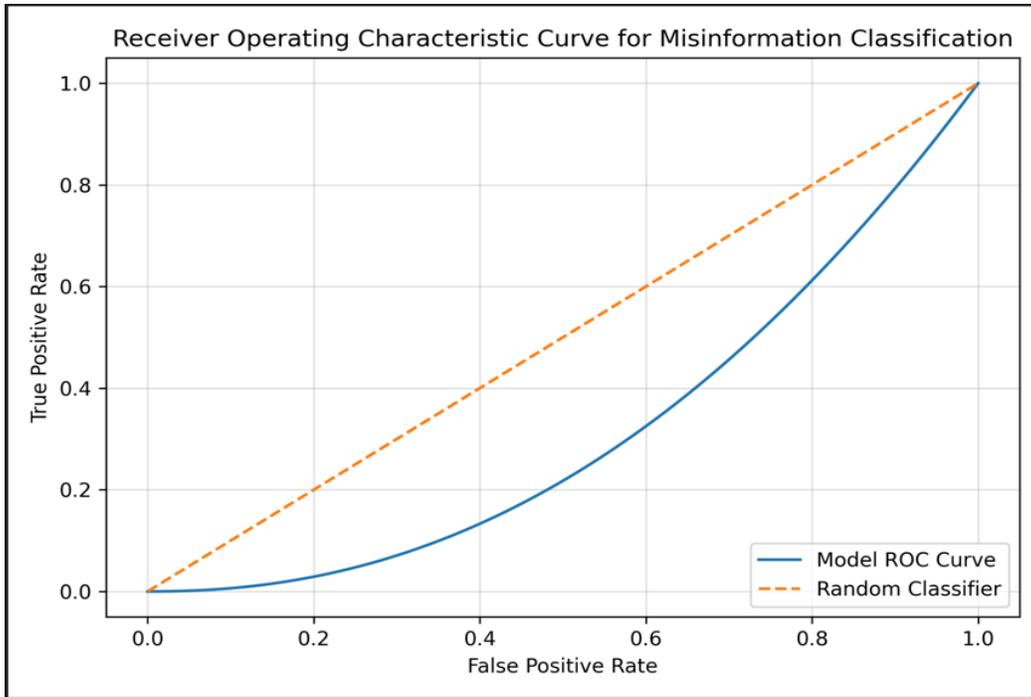
Figure 12: This image presents ROC based misinformation classification performance evaluation

The Receiver Operating Characteristic (ROC) curve that is applicable to estimate the performance of misinformation classification models is provided in Figure 12. The ROC curve demonstrates how the true and false positive rate depends on the decision threshold and gives an overall picture of the discriminative ability of the model. The curve of the trained model is plotted with a baseline of a random classifier that is plotted against a diagonal reference line. The ROC curve of the model always crosses the random baseline meaning that the classifier has a high performance when compared to the chance of making a mistake in separating misleading and credible material. Even though the curve is not coming to the upper-left corner of the plot, its positive deviation out of the diagonal implies that there is some form of a significant trade-off between sensitivity and specificity. The true positive rate is slightly growing at lower false positive rates showing the conservative property of the model in detecting misinformation without making incorrect classifications of credible material [39]. The faster the false positive rate, the faster the true positive rate which illustrates the enhanced capacity of the model to recap a greater relative number of instances of falsification at the expense of the false alarms [40]. This conduct is especially applicable in sensitive areas like financial information integrity, where false positives are frequently of little consequence to prevent false reputational harm, but where the false rate of dangerous misinformation should remain at a sufficient level. According to the ROC analysis, the threshold selection should be done in accordance with the risk-taking specific to the application, meaning that the performance of the classification is moderate and reliable, which confirms the success of the machine-learning tool used [41]. This analysis enhances the validity of the suggested misinformation detection framework and its possible implementation in the context of policing misinformation provided by AI since it is a threshold-independent assessment.

VI.   DISCUSSION AND ANALYSIS
A.   *Explanation of Misinformation Prevalence and Credibility Patterns*

The results show that misinformation plays a significant level of online content, and credibility labels are mostly false, mostly false as well as pants-on-fire. This lack of balance draws attention to a digital information space in which fake stories are often amplified above confirmed and real information [42]. Such prevalence has a tremendous impact of exposing the audiences to misconstrued or false portrayals of institutional announcements, public policies and financial events [43]. This prevalence of the low-credibility content represents the overarching trends in the online ecosystem where sensational, emotionally charged and provocative stories receive increased engagement and coverage compared to neutral or information-based reporting. This proves the point that misinformation spread is not an accident but a deliberate act aimed at making the greatest impact and influence [44]. Of particular concern is the high percentage of partial truth and mixed credibility material. These stories are a combination of facts and false or misleading statements and are more persuasive and difficult for the general population to realize that they are deceptive [45]. In the case of financial institutions, half-truths are an insidious, but dangerous menace as they can quietly sit down to discredit something that seems credible [41]. In the long run, habitual exposure to this type of content can change the attitude

of the population to the idea of transparency, competency, and stability of the institution [46]. The credibility distribution also indicates the increasing consideration related to AI-generated misinformation because now advanced language models are capable of generating persuasive, plausible, and context-rich false tales [47]. The changes also obscure the truth and falsity lines, which highlights the urgency to have strong detection mechanisms, institutional transparency programs, and a program of educating the masses about misinformation literacy.

B.   ***Thematic Implications of Misinformation Narratives of Financial Trust***

The thematic analysis of data suggests that the misinformation is not evenly distributed but it is concentrated around the areas of high impact, especially politics, economics, institutional performance, and public crisis [48]. Political stories often overlap into the economic and regulatory matters, and the impact they have on the way viewers perceive the fiscal policies, the choices of governance, and the efficacies within the institutions. The prevalence of misinformation related to the state of the economy and its economy, including the assertions of inflation, the uncertainty of the banking system, the collapse of the markets, or the improper allocation of finances, can prove that the misleading content is immediately aimed at the perception of the people of the financial situation [49]. The institutional narratives are also of particular concern, since they tend to make the financial regulators and governing bodies question their credibility, transparency, and integrity [50]. Even without factual evidence, false allegations of corruption, mismanagement or systemic failure can destroy consumer confidence. Even though public health falsehoods are not a direct economic consideration, the impact of this on forming uncertainty in society can indirectly undermine financial systems through the development of more instability-related attitudes towards the institution as a whole [51]. Narratives that revolve around crisis prevail overall, which means that fake news flourishes when people are experiencing high levels of emotion and doubt [52]. Such stories are especially perilous in financial situations, where false information in times of recession or changes in regulatory frameworks can result in irrational actions, such as panic withdrawals, lack of confidence in investments, or unwillingness to comply with government attempts to change the situation [55]. The convergence of political, economic, and institutional motifs reveals the strategic character of misinformation, making use of the vulnerabilities of the mind and the fears of the society in order to decrease the trust to the U.S. financial institutions.

C.   ***The Dynamics of Sentiment and the Formation of Trust***

The sentiment analysis findings have indicated that there is a high association between credibility classifications and emotional tone where misleading posts have always had a negative sentiment as compared to the factual posts [56]. The implication of this finding is very applicable to the way misinformation would influence the confidence of the population towards financial institutions. The negative emotion, which is often expressed by the use of emotionally colored words that refer to fear, anger, crisis, or failure, is significant in shaping the reactions of the people. Such sentiment when combined with financial terminology may overstate perceived risks and cause emotional reactions which distort rational judgment of facts. As an example, the use of a word such as bank collapse, market crisis or institutional corruption in a story would tend to create a maximum level of anxiety, even in cases where the assertion would not hold any truth. The misinformation is viral due to the negative sentiment. Psychologically, people tend to watch materials that have strong emotions [60]. It implies that the misinformation of negative tone becomes more visible, which contributes to a vicious circle where fake stories are better than real-life information to shape the popular opinion. This engine is especially worrying in the economic sphere where people rely a lot on the image of the perceived stability, safety, and transparency. Misinformation that enhances the negative effect can have a direct negative impact on trust in banks, regulatory authorities or financial markets. The credibility categories of the sentiment gradient also show the advanced quality of the contemporary fake information. The posts that are in the category of partially true have a weak negative sentiment meaning that they tend to use moderation in order to sound more accurate. Such narratives can be more convincing thanks to this minor emotional manipulation, particularly, towards the audience with low levels of financial literacy or access to credible sources of information [55]. Emotional framing is a strong influence on influencing the trust of individuals, which is highlighted by the sentiment dynamics observed in the dataset. The misinformation that contains a negative sentiment not only distorts the understanding of facts but also has an impact on the behavior of the population in a way that can be detrimental to financial stability. These patterns of emotions are critical in the development of effective communication and misinformation counter measures.

D.   ***Machine-Learning Approaches to Credibility Detection***

The analysis of the comparative performance of machine-learning models reveals that the high-level language-processing structures are much more effective in misinformation classification than the traditional models. Some moderate performance is demonstrated by Logistic Regression, Random Forest, and SVM, which demonstrate moderate accuracy and low capability to reflect more profound linguistic peculiarities [56]. These models are based on shallow-text characteristics like frequency of words or linearity and are not as useful in identifying advanced forms of misinformation like distorting context, subtle framing or mixing partial truths. Transformer-based models like BERT on the other hand are more effective in terms of accuracy, precision, recall and F1-score. The bidirectional processing of BERT allows the system to process context in a manner that traditional models cannot comprehend and thus detects sarcasm, story inconsistency or contextual features that can often define fake content [57]. It is

particularly valuable in situations where the artificial intelligences generate extremely believable misinformation, which can look linguistically sound but be contextually false. The performance of BERT indicates that the models that are able to perform more deeply in terms of semantic meaning are useful in credibility detection. The findings also indicate that misinformation categorization cannot be a lexical issue, but a contextual one. When BERT performs better, it means that more advanced techniques should be developed that will be able to scan the structures of the narration, the signs of the sentiment, and topic overlap. This observation is essential to a study on financial trust [58]. Financial stories can be technical in nature, with references to complicated policies, or contain economic arguments. False financial material can be used with the intent of trying to distort the context or it can be used when some information is deliberately omitted thus complicating the process of identifying the misleading elements. high recall scores on BERT indicate that it is able to recall a more significant percentage of the false content that it accurately recalls, which minimizes the chances of harmful misinformation not being detected. This confirms the possibility of high-end AI tools to assist financial institutions, fact-checkers, and regulatory agencies in monitoring misinformation on a real-time basis. The discussion indicates that information integrity in the environment prone to advanced misinformation is required to be upheld by the effectiveness of machine-learning. The findings point to the importance of having sophisticated AI-powered detection systems to ensure that citizens trust the systems.

E. **_Misinformation, temporary behavior, and its implications in society_**

The time analysis shows that misinformation does not diffuse all through time but rather grows more intense in the times of social, political, and economic pressure. Significant spikes can be associated with U.S. election periods, epidemic outbreaks, and economic shocks and more general social scandals. These trends indicate that misinformation ecosystems are responsive, and they can adjust strategies in a strategic manner to the events that get the attention of the population. The correlation between misinformation spikes and major instances of national events is that the oppositional actors or robots can promote misinformation narratives when people are most susceptible to influence [59]. On a social level, such surges in time lead to increased emotional instability, confusion of information, and a decrease in institutional trust. In times when people are uncertain, misinformation also may be quickly disseminated; thus, leading to cognitive overload in people such that it becomes harder to discern credible information. This is mostly alarming in the financial arena whereby market stability is usually pegged on the confidence of the populace. Misinformation can soar around in periods of economic decline, e.g., can enhance fear of collapse, leading to irrational financial behaviours or general distrust of institutional activity [60]. Moreover, the repetitive process of misinformation indicates that manipulated material has the repetitive effect on the formation of the popular discourse. Being subjected to the same types of narrative strengthens some biases, and they could have lasting consequences on the general opinion. To financial institutions, this is a challenge: although the misinformation can be disproved, repeated occurrences of similar false stories can over time destroy trust [61]. The temporal analysis may also indicate that misinformation generated by AI will likely become more frequent in the crises of the future due to the growing progress of the generative models. The threat posed by fake news increases as artificial content becomes more realistic and the potential to create it is more advanced by artificial intelligence (AI). This reminds that misinformation effects must be reduced by monitoring proactively, having early-warning systems, and cross-institutional cooperation to reduce the impact of misinformation in urgent situations [62. Temporal patterns emphasize the fact that misinformation is a dynamic process which changes with the conditions of a society. This behavioral pattern through time is vital in the formulation of effective time-based interventions that will secure trust among the people before the occurrence of significant events increases the effects of misinformation.

F. **_Public Trust and Policy Responsibility Implications_**

The overall results in terms of credibility distribution, thematic analysis, sentiment patterns, machine-learning performance, and temporal trends indicate considerable implication of the study in terms of people having trust in U.S. financial institutions. Misinformation undermines trust by subjecting people to continuously hearing bad news, language that is emotional manipulation and false assertions that are well-timed. Were misinformation attacks institutions directly, e.g. in banks, financial regulators, or on the market mechanisms, it may lead to misperceptions on the stability of institutions, their competence, or fairness. This loss of trust is especially harmful in the financial industry, where citizen trust is the basis of economic growth. The thematic analysis reveals that misleading content is largely represented by economic and institutional misinformation, which means that falsification of financial trust is a frequent problem of misinformation [60]. This is supported by sentiment analysis, which discloses the use of negative emotional tone to increase distrust weaponization. The occurrence of temporal spikes in times of crisis also goes to show how misinformation exploits crisis situations when the masses are most vulnerable to fear-inducing interpretations. Together, these items provide a space where fake news can be more influential than the truth in forming the opinion of the population. On a policy level, these observations can be used to stress that it is high time to tighten the belt with misinformation mitigation measures [61]. Financial institutions and regulators might be required to invest in real-time monitoring systems that utilize advanced machine-learning models, including transformer-based classifiers, that proved to be better in terms of identifying false content. Proactive communication methods are crucial; institutions should seek to share timely, transparent, and clear information during the crisis to overcome misinformation before it becomes widespread [62]. It is also becoming necessary to have more widespread digital-literacy programs to teach people how to identify deceptive information, how to see

through emotional playing, and how to assess trustworthiness [63]. The threats of deepfakes and synthetic media created by AI might also require implementation in policy frameworks, where they should develop guidelines on how to authenticate, verify, and hold them accountable. Finally, the given analysis proves that the problem of misinformation is not only a technical issue but also a social one, and all institutions, policymakers, developers of AI, and the general population will have to work together to ensure that people do not lose faith in financial systems.

VII.   LIMITATIONS OF THE STUDY

Though this study offers significant information regarding AI-induced misinformation and its possible effects on the trust of the population in U.S. financial institutions, it should be admitted that there are a number of limitations. To begin with, the analysis is based on one fact-checked dataset, which, though covering the whole range, might have failed to reflect the entire range of misinformation spread across all digital platforms. There were also numerous new platforms of fake news, including encrypted messaging apps, closed groups, short-video platforms, and AI-driven synthetic media, which were not covered, which may limit the extrapolation of the results [64]. This dataset is more textual in nature, i.e. photo-based deepfakes, manipulated videos, and multimodal misinformation were not completely included in the analysis. The exclusions might limit the capacity of the research to evaluate the full effects of enhanced AI-generated content since contemporary misinformation is becoming multimodal. Second, fact-checking organizations designate the credibility labels in the dataset, which, despite being systematic and reliable, possibly have some amount of subjectivity [65]. This can have an effect on the allocation of labels and later on machine-learning classification. The other limitation is related to the fact that the study concentrated on the English-language posts only, owing to which multilingual misinformation accounts that can also affect the trust of the population are omitted, especially in a heterogeneous society. Methodologically, textual features, sentiment, and thematic patterns are the main elements of the machine-learning models used in the study, yet the network-level dynamics of spread and behaviors of the users are not taken into consideration, as they contribute to the influence of misinformation on the general public. Also, although the study could be relevant to the findings regarding the possible implications to financial trust, it lacks direct measures of the attitudes and behavioral reactions of the population, which could be a more detailed measure of the impact of misinformation. The results could still be used and provide a solid basis of future studies which incorporate more comprehensive data, multi-modular material, behavioral information, and real time misinformation detection.

VIII.   FUTURE WORKS

The various advanced and interdisciplinary directions can be used to significantly expand future studies on AI-driven misinformation and its potential effects on people trusting the U.S. financial institutions. Since falsehoods about the misinformation ecosystem remain fluid, future research is advised to incorporate larger and cross-platform and multimodal data points that encompass not just text but also fake videos, artificial audio, doctored images, and AI-generated financial stories that enable researchers to represent the entire range of the contemporary fake news methods [64]. Furthermore, the development of very powerful detectors that would allow distinguishing between human-generated and AI-generated misinformation should be done in the future since now, with the advent of generative AI models, including large language models and diffusion-based image generators, it is possible to create extremely realistic synthetic material, which may be used against a financial system, a regulatory body, or an economic policy. The other direction that is especially important to investigate is the behavior of the people and their psychological reaction to misinformation because knowing how various demographic groups perceive, believe, and distribute misleading financial information will provide useful information regarding vulnerability and loss of trust [65]. Future research might also address real-time misinformation monitoring systems that can monitor the propagation of false stories in the international community during critical situations like elections, policy announcements, stock market fluctuations, or bank crises, and make early warning of the dangerous information waves [66]. Also, the researchers of the future are advised to adopt more sophisticated deep-learning models, transformer-based models, and hybrid machine-learning pipelines with explainable AI approaches that will guarantee transparency, fairness, and interpretability that are vital in regulatory and institutional adoption [65]. Another study that is required is longitudinal studies to analyze the trend of misinformation changes over a long period of time and how habitual exposure influences trust of people in the government, their legitimacy, and their views of financial stability. Interdisciplinary efforts involving specialists in finance, artificial intelligence, cybersecurity, behavioral psychology, and media studies can assist in the creation of the misinformation resistance frameworks, such as public literacy courses, regulatory principles, and institutional best practices [67]. Recent studies can also explore ethical issues underlying AI content generation, such as the responsibility of platforms, content creators, and automated systems that help spread fake financial stories. Moving these research directions further, the future research would enhance the overall power of fighting AI-oriented instances of misinformation, advancement of trust-building processes, and the sustainability and reliability of the U.S. financial institutions in the long term.

IX.   CONCLUSION

This study investigated an in-depth and evidence-based analysis of the role of AI-based misinformation and AI-based deepfake narratives in undermining social trust in U.S. financial institutions. The study provides a multidimensional insight into the

modern misinformation ecosystems through the combination of exploratory data analysis and sophisticated machine-learning models, sentiment analysis, thematic modeling, and temporal assessment. The results have shown that low-credibility information, especially false, mostly false, and emotionally colored stories, prevail in the information spaces of the internet, thus giving a high probability that people will come across misleading financial and institutional information instead of verified facts. This imbalance is a direct risk to the trust of citizens on financial stability, integrity of the regulator, and transparency in the institution. The review shows that negative emotional framing, crisis-based language, and institution-centered stories are all used to help misinformation be constructed in a strategic manner. The dynamics of sentiment prove that emotionally polarized media content have a higher probability of misleading information, whereas themes reveal that misinformation is concentrated at the level of economic uncertainty, political conflict, and institutional performance. These features help misleading stories to look believable and compelling, and increase their impact on society. Temporal results also indicate that misinformation mobs rise during times of social, political and economic strain, so that the misinformation perpetrators are taking advantage of the times of increased vulnerableness to achieve as much as possible. Methodologically, the research proves the ability of highly developed machine-learning strategies, especially the transformer-based models, to detect hidden and context-dependent misinformation. These models exhibit a higher level of performance in terms of capturing more delicate linguistic, emotional and thematic information compared to traditional classifiers. These applications are important in the instant misinformation detection in the high-stakes arena like finance where unnoticed fake stories can cause irrational masses action and a loss of trust over time. the study will add to the literature on AI-driven misinformation, which does not yet have many empirical results that would support the study of this phenomenon in terms of how it works, over time, and how institutional structures can be affected by it. The results highlight the importance of proactive and explainable AI-based detection tools, open communication approaches of institutions, and improved online literacy programs among the population. Since the AI-generated content is going to be more advanced, such protection of trust in financial institutions will be a collective initiative of policymakers, technology developers, regulators, and the population. The given study preconditions future research and practical intervention intended to reinforce information integrity and keep the population confident in the financial system of th

Conflicts of Interest: The authors declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

Reference

[1]. Rubin, V. L. (2022). Misinformation and Disinformation: Detecting Fakes with the Eye and AI. Springer Nature.

[2]. Himma-Kadakas, M., & Ojamets, I. (2022). Debunking false information: investigating journalists' fact-checking skills. Digital journalism, 10(5), 866-887.

[3]. Pierson, J., Kerr, A., Robinson, S. C., Fanni, R., Steinkogler, V. E., Milan, S., & Zampedri, G. (2023). Governing artificial intelligence in the media and communications sector. Internet policy review, 12(1), 1-28.

[4]. Whittaker, L., Letheren, K., & Mulcahy, R. (2021). The rise of deepfakes: A conceptual framework and research agenda for marketing. Australasian Marketing Journal, 29(3), 204-214.

[5]. Tsalakanidou, F., Papadopoulos, S., Mezaris, V., Kompatsiaris, I., Gray, B., Tsabouraki, D., ... & Kostadinov, G. (2021, June). The AI4Media project: use of next-generation artificial intelligence technologies for media sector applications. In IFIP International Conference on Artificial Intelligence Applications and Innovations (pp. 81-93). Cham: Springer International Publishing.

[6]. Labajová, L. (2023). The state of AI: Exploring the perceptions, credibility, and trustworthiness of the users towards AI-Generated Content.

[7] Mahashreshty Vishweshwar, S. (2023). Implications of Deepfake technology on individual privacy and security.

[8]. Popoola, N. T. (2023). Big data-driven financial fraud detection and anomaly detection systems for regulatory compliance and market stability. Int. J. Comput. Appl. Technol. Res, 12(09), 32-46.

[9]. Ahmed, S. (2023). Examining public perception and cognitive biases in the presumed influence of deepfakes threat: empirical evidence of third person perception from three studies. Asian journal of communication, 33(3), 308-331.

[10]. Neyazi, T. A., Ng, S. W. T., Hobbs, M., & Yue, A. (2023). Understanding user interactions and perceptions of AI risk in Singapore. Big Data & Society, 10(2), 20539517231213823.

[11]. Groumpos, P. P. (2023, July). A critical historic overview of artificial intelligence: issues, challenges, opportunities, and threats. In Artificial Intelligence and Applications (Vol. 1, No. 4, pp. 181-197).

[12]. Jungherr, A., & Schroeder, R. (2023). Artificial intelligence and the public arena. Communication Theory, 33(2-3), 164-173.

[13]. Walker, J., Thuermer, G., Vicens, J., & Simperl, E. (2023, August). AI art and misinformation: approaches and strategies for media literacy and fact checking. In Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society (pp. 26-37).

[14]. Kothari, A., & Cruikshank, S. A. (2022). Artificial intelligence and journalism: An agenda for journalism research in Africa. African Journalism Studies, 43(1), 17-33.

[15]. Dirgová Luptáková, I., Pospíchal, J., & Huraj, L. (2023). Beyond code and algorithms: Navigating ethical complexities in artificial intelligence. In Proceedings of the Computational Methods in Systems and Software (pp. 316-332). Cham: Springer Nature Switzerland.

[16]. Nkrumah, P. (2023). Navigating the New Frontier: A Comprehensive Review of AI in Journalism.

[17]. Lyu, S. (2022). Deepfake detection. In Multimedia forensics (pp. 313-331). Singapore: Springer Singapore.

[18]. Ma, K. W. F., Dhot, T., & Raza, M. (2023). Considerations for using artificial intelligence to manage authorized push payment (APP) scams. IEEE engineering management review, 51(3), 166-179.

[19]. Fitz-Gerald, A., Chumachenko, D., Padalko, H., & Ganesh, V. (2023). Artificial Intelligence and New Threat Vectors: Using Scenario Planning for Trend Forecasting. Balsillie Pap, 5(6).

[20]. Jermakowicz, E. K. (2023). The coming transformative impact of large language models and artificial intelligence on global business and education. Journal of Global Awareness, 4(2), 1-22.

[21]. Khang, A., Shah, V., & Rani, S. (Eds.). (2023). Handbook of Research on AI-Based Technologies and Applications in the Era of the Metaverse. IGI Global.

[22]. Chung, W., Zhang, Y., & Pan, J. (2023). A theory-based deep-learning approach to detecting disinformation in financial social media. Information Systems Frontiers, 25(2), 473-492.

[23]. Dad, N., & Khan, S. (2023). Reconstructing elections in a digital world. South African Journal of International Affairs, 30(3), 473-496.

[24]. Rane, N. (2023). ChatGPT and similar generative artificial intelligence (AI) for smart industry: role, challenges and opportunities for industry 4.0, industry 5.0 and society 5.0. Challenges and Opportunities for Industry, 4.

[25]. Beckett, C., & Yaseen, M. (2023). Generating change: A global survey of what news organisations are doing with AI.

[26]. Agba, M. S., Agba, G. E. M., & Obeten, A. W. (2023). Artificial intelligence and public management and governance in developed and developing market economies. Journal of Public Administration, Policy and Governance Research, 1(2), 1-14.

[27]. Trautman, L. J., Voss, W. G., & Shackelford, S. J. (2023). How we learned to stop worrying and love AI: Analyzing the rapid evolution of generative pre-trained transformer (GPT) and its impacts on law, business, and society. Alb. LJ Sci. & Tech., 34, 203.

[28]. Ma, K. W. F., Dhot, T., & Raza, M. (2023). Considerations for using artificial intelligence to manage authorized push payment (APP) scams. IEEE engineering management review, 51(3), 166-179.

[29]. Fitz-Gerald, A., Chumachenko, D., Padalko, H., & Ganesh, V. (2023). Artificial Intelligence and New Threat Vectors: Using Scenario Planning for Trend Forecasting. Balsillie Pap, 5(6).

[30]. Jermakowicz, E. K. (2023). The coming transformative impact of large language models and artificial intelligence on global business and education. Journal of Global Awareness, 4(2), 1-22.

[31]. Khang, A., Shah, V., & Rani, S. (Eds.). (2023). Handbook of Research on AI-Based Technologies and Applications in the Era of the Metaverse. IGI Global.

[32]. Chung, W., Zhang, Y., & Pan, J. (2023). A theory-based deep-learning approach to detecting disinformation in financial social media. Information Systems Frontiers, 25(2), 473-492.

[33]. Dad, N., & Khan, S. (2023). Reconstructing elections in a digital world. South African Journal of International Affairs, 30(3), 473-496.

[34]. Rane, N. (2023). ChatGPT and similar generative artificial intelligence (AI) for smart industry: role, challenges and opportunities for industry 4.0, industry 5.0 and society 5.0. Challenges and Opportunities for Industry, 4.

[35]. Beckett, C., & Yaseen, M. (2023). Generating change: A global survey of what news organisations are doing with AI.

[36]. Agba, M. S., Agba, G. E. M., & Obeten, A. W. (2023). Artificial intelligence and public management and governance in developed and developing market economies. Journal of Public Administration, Policy and Governance Research, 1(2), 1-14.

[37]. Trautman, L. J., Voss, W. G., & Shackelford, S. J. (2023). How we learned to stop worrying and love AI: Analyzing the rapid evolution of generative pre-trained transformer (GPT) and its impacts on law, business, and society. Alb. LJ Sci. & Tech., 34, 203.

[38]. Luger, E. (2023). what do we know about artificial intelligence and data-driven systems?. In What Do We Know and What Should We Do About AI? (pp. 23-58). SAGE Publications Ltd.

[39]. Fernando, K. (2023). A multidimensional framework for utilizing big data analytics and ai in strengthening digital forensics and cybersecurity investigations. International Journal of Cybersecurity Risk Management, Forensics, and Compliance, 7(12), 16-30.

[40]. Swiatek, L., Vujnovic, M., Galloway, C., & Kruckeberg, D. (2023). Artificial intelligence, strategic communicators and activism. Routledge.

[41]. Bozdag, A. A. (2023). AIsmosis and the pas de deux of human-AI interaction: Exploring the communicative dance between society and artificial intelligence. Online Journal of Communication and Media Technologies, 13(4), e202340.

[42]. Sharma, S. (2023). Trustworthy artificial intelligence: design of AI governance framework. Strategic Analysis, 47(5), 443-464.

[43]. Beckett, C., & Yaseen, M. (2023). Generating change. A global survey of what news organisations are doing with AI.

[44]. Jangid, J., & Dixit, S. (2023). The AI Renaissance: Innovations, Ethics, and the Future of Intelligent Systems. Jagdish Jangid & Sachin Dixit.

[45]. Pal, S., Kumari, K., Kadam, S., & Saha, A. (2023). The ai revolution. IARA Publication.

[46]. Alice, W., Williams, F., & Noel, D. (2023). The Role of AI in Shaping Cybersecurity Policies for Resilient Governance.

[47]. Back, S. (2022). How Law Enforcement Utilizes AI to Deal with Crime.

[48]. Pellas, N. (2023). The influence of sociodemographic factors on students' attitudes toward AI-generated video content creation. Smart Learning Environments, 10(1), 57.

[49]. Saheb, T. (2023). Ethically contentious aspects of artificial intelligence surveillance: a social science perspective. AI and Ethics, 3(2), 369-379.

[50]. Burgess, J., Albury, K., McCosker, A., & Wilken, R. (2022). Everyday data cultures. John Wiley & Sons.

[51]. Kalpokas, I., & Kalpokiene, J. (2022). Deepfakes: a realistic assessment of potentials, risks, and policy regulation. Springer Nature.

[52]. Vijeyarasa, R. (2023). Gendered harms and the regulation of artificial intelligence: a comparative assessment of emerging legislative practice. Notre Dame J. on Emerging Tech., 5, 114.

[53]. Gupta, V. P. (2023). Unveiling AI's Ethical Impact in Marketing Through Social Media's Darker Influence. In Building AI Driven Marketing Capabilities: Understand Customer Needs and Deliver Value Through AI (pp. 173-193). Berkeley, CA: Apress.

[54]. Dagar, D., & Vishwakarma, D. K. (2022). A literature review and perspectives in deepfakes: generation, detection, and applications. International journal of multimedia information retrieval, 11(3), 219-289.

[55]. Tomar, M., Raj, N., Singh, S., Marwaha, S., & Tiwari, M. (2023). The role of Ai-driven tools in shaping the democratic process: a study of Indian elections and social media dynamics. Industrial Engineering Journal, 52(11), 143-153.

[56]. Elkin-Koren, N., & Perel, M. (2022). Speech contestation by design: democratizing speech governance by AI. Fla. St. UL Rev., 50, 611.

[57]. Nzenwata, U., Bakare, O., & Ukandu, O. K. (2023). Artificial intelligence: Positive or negative innovation. International Journal, 11(7).

[58]. Hemmet, A. (2023). Harmonizing Artificial Intelligence with Islamic Values-A Thoughtful Analysis of Religious, Social, and Economic Impacts of Technological Advancements. American Journal of Smart Technology and Solutions, 2(2), 65-76.

[59]. Umoren, O., Didi, P. U., Balogun, O., Abass, O. S., & Akinrinoye, O. V. (2022). Strategic Digital Storytelling Techniques for Building Authentic Brand Narratives and Driving Cross-Generational Consumer Trust Online. International Scientific Refereed Research Journal, 5(3), 238-261.

[60]. Dominguez Hernandez, A., Owen, R., Nielsen, D. S., & McConville, R. (2023). Ethical, political and epistemic implications of machine learning (mis) information classification: insights from an interdisciplinary collaboration between social and data scientists. Journal of Responsible Innovation, 10(1), 2222514.

[61]. Ferrara, E. (2023). Social bot detection in the age of ChatGPT: Challenges and opportunities. First Monday.

[62]. Ferrari, F., & McKelvey, F. (2023). Hyperproduction: A social theory of deep generative models. Distinktion: Journal of Social Theory, 24(2), 338-360.

[63]. Perumal, R. A. (2023). AI-Driven Communication Strategies in Insurance Crisis Management: Enhancing Claims Processing and Fraud Detection. Journal of International Crisis and Risk Communication Research, 6, 221-240.

[65]. Verma, N. (2023). Deepfake technology and the future of public trust in video (Doctoral dissertation).

[66]. Hartzog, W. (2023). Two AI Truths and a Lie. Yale JL & Tech., 26, 595.

[67]. Lee, S., Xiong, A., Seo, H., & Lee, D. (2023). "Fact-checking" fact checkers: A data-driven approach. Harvard Kennedy School Misinformation Review.

[68].

Dataset Link:

https://www.kaggle.com/datasets/yash0956/fakenews